

BAYESIAN ANALYSIS OF SMALL DOMAIN DATA IN REPEATED SURVEYS

By

NARINDER K. NANGIA

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL  
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT  
OF THE REQUIREMENTS FOR THE DEGREE OF  
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

1992

## ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to Professor Malay Ghosh for his invaluable guidance and support throughout this project. I would like to thank Dr. Ronald H. Randles, Dr. P. V. Rao, Dr. Ramon C. Littell, Dr. Geoffrey G. Vining, and Dr. Sencer Yeralan for being on my committee.

Much gratitude is owed to my parents, whose support, advice, guidance and prayers throughout the years of my life have made this achievement possible. My debt to them can never be repaid in full. A very special thank you is offered to my wife, Uma, for her continuous comfort, trust, and understanding; and my son, Nishant, for his loving support.

## TABLE OF CONTENTS

|                                                                  |    |
|------------------------------------------------------------------|----|
| ACKNOWLEDGEMENTS . . . . .                                       | ii |
| ABSTRACT . . . . .                                               | v  |
| CHAPTERS                                                         |    |
| 1 INTRODUCTION . . . . .                                         | 1  |
| 1.1 Literature Review . . . . .                                  | 1  |
| 1.2 The Subject of This Dissertation . . . . .                   | 4  |
| 2 ESTIMATION OF SMALL AREA MEANS IN TIME DOMAIN . . .            | 6  |
| 2.1 Introduction . . . . .                                       | 6  |
| 2.2 Description of the Hierarchical Bayes Model . . . . .        | 8  |
| 2.3 Empirical Bayes Estimation . . . . .                         | 14 |
| 2.4 An Application of Hierarchical Bayes Analysis . . . . .      | 16 |
| 3 MULTIVARIATE ESTIMATION OF MEANS IN TIME DOMAIN .              | 30 |
| 3.1 Introduction . . . . .                                       | 30 |
| 3.2 The General Multivariate Hierarchical Bayes Model . . . . .  | 31 |
| 3.3 An Illustrative Application to Survey data . . . . .         | 33 |
| 4 HIERARCHICAL BAYES PREDICTION IN FINITE POPULATIONS            | 46 |
| 4.1 Introduction . . . . .                                       | 46 |
| 4.2 Description of the Hierarchical Bayes Model . . . . .        | 48 |
| 4.3 The Hierarchical Bayes Predictor in a Special Case . . . . . | 57 |
| 4.4 Best Unbiased Prediction in Small Area Estimation . . . . .  | 59 |
| 4.5 Stochastic Domination . . . . .                              | 67 |
| 4.6 Best Equivariant Prediction . . . . .                        | 72 |
| 5 SUMMARY AND FUTURE RESEARCH . . . . .                          | 86 |
| 5.1 Summary . . . . .                                            | 86 |

|                               |                                  |    |
|-------------------------------|----------------------------------|----|
| 5.2                           | Future Research . . . . .        | 87 |
| APPENDIX                      | PROOF OF THEOREM 4.2.1 . . . . . | 88 |
| BIBLIOGRAPHY . . . . .        |                                  | 94 |
| BIOGRAPHICAL SKETCH . . . . . |                                  | 97 |

Abstract of Dissertation Presented to the Graduate School  
of the University of Florida in Partial Fulfillment  
of the Requirements for the Degree of  
Doctor of Philosophy

## BAYESIAN ANALYSIS OF SMALL DOMAIN DATA IN REPEATED SURVEYS

By

NARINDER K. NANGIA

December 1992

Chairman: Malay Ghosh  
Major Department: Statistics

We consider in this dissertation estimation and prediction of local area means based on repeated surveys over time. A general hierarchical Bayes model is used for this purpose. The resulting HB estimators for a small area "borrow strength" over time in addition to borrowing strength from other local areas. The Bayesian procedure provides current as well as retrospective estimates simultaneous for all the small areas. The HB methodology, in infinite as well as finite population sampling situations, has been implemented by adopting a Monto Carlo integration technique—the Gibbs Sampler.

A multivariate HB model is considered to exploit the inter-relationship among the parameters in improving the small area estimators. Both the univariate and multivariate methodologies have been implemented on median income estimation problem. The results show that the estimators borrowing strength over time and from other local areas are substantially better than the estimators which borrow strength only from other local areas.

In finite population sampling, the predictive distribution of a characteristic of interest for the unsampled population units is found given the observations on the sampled units and is used to draw inference. In particular, we have provided HB predictors and the associated standard errors of the local area means over time. In a special case of this HB analysis, when the vector of the ratios of the variance components is known, the HB predictor of the vector of local area means is shown to possess some frequentist optimality properties. For example, among other properties, we have shown that the HB predictors are best unbiased, best equivariant predictor under a suitable group of transformations.



## CHAPTER 1

### INTRODUCTION

#### 1.1 Literature Review

The existence of small area (domain) statistics based either on a census or on other administrative records aiming at complete enumeration can be traced back to 11th Century in England and 17th Century in Canada (see Brackstone (1987)). For the past few decades, sample surveys for most purposes have taken the place of complete enumeration as means of data collection. These surveys have been able to provide sample based estimates of adequate precision of the parameters of interest at the national and subnational levels. However, the overall sample size in such surveys is usually determined with the target of providing specified accuracy at a much higher level of aggregation than that of the small areas. The resulting small sample sizes for the small areas are usually insufficient to provide sample based estimates of adequate precision of parameters at the small area level and are likely to lead to large estimation errors. On the other hand, the reliability of the small area statistics is crucial for regional economic planning and administration. The evaluation of state economic policies and identification of the subgroups or regions of the population which are not keeping up with average, relies heavily on the small area statistics. Furthermore, the decisions concerning the investments and the location of small industrial units by the corporate sector are usually based on the local social, economic, and environmental

conditions. As a consequence, there is a growing interest in the development of reliable small area statistics.

In recent years, several model-based estimation techniques have emerged resulting in improved small area estimates. The essence of these small area estimation techniques is to borrow strength from similar neighboring areas for estimation and prediction purposes. These estimation techniques use models, either explicitly or implicitly, that "connect" the small areas through supplementary data (e.g., census and administration data). Early work of Ericksen (1973, 1974) employed the simple regression method that combines sample data with symptomatic and census data for estimating the population of local areas. By adopting a regression model similar to that of Ericksen (1974), Fay and Herriot (1979) suggested a procedure that used a constraint weighted average of the model-dependent and sample-based estimators with weights inversely proportional to the average lack of fit of the regression, and the variance of the sample estimate. These weighted average estimates were constrained to be within one standard error of the sample estimates. Ghosh and Rao (1991) have recently surveyed most of the existing small area estimation methods.

The simultaneous estimation of the parameters at a given time for several small areas (strata) by employing a general linear model with covariates have been considered by Fay and Herriot (1979), Ericksen and Kadane (1985, 1987), Ghosh and Meeden (1986), Ghosh and Lahiri (1988), Battese, Harter and Fuller (1988), Prasad and Rao (1990) using either a variance components approach or an empirical Bayes (EB) approach. An EB approach does not incorporate uncertainty due to the estimation of prior parameters and thus leads to underestimation of true posterior variance. The underestimation problem gets more severe when one or more variance components of the model are unknown.

An alternative hierarchical Bayes (HB) procedure for the small area estimation at a given time, has been considered in Ghosh and Lahiri (1988) and Datta and Ghosh



(1991). Both the EB and IIB procedures often lead to similar point estimates of the parameters of interest. We may also add that in contrast to the EB approach, the HB approach provides small area estimators that have natural measures of standard errors associated with them, namely the posterior standard deviations.

In the sample survey literature, the importance of model-based approach to small area estimation has been well recognized. However, the model-based techniques for small area estimation considered so far borrowed strength only from similar other areas to improve small area estimators.

However, often these small area (strata) populations are subject to change in time. Usually the time series information is available through the repeated surveys, (e.g. market research survey, crop cutting experiments etc.) carried out at regular time intervals by industries and government agencies. As mentioned in Scott and Smith (1974), Indian National Sample Survey, a multipurpose survey carried out annually since 1950 and the Medical Data Index in The U.K., a market research survey of doctors are examples of nonoverlapping surveys with human populations. In such surveys, one has at one's disposal not only the current data, but also data from similar past experiments. Typically, only a few samples are available from an individual area.

In the past, time series models have usefully been applied to the analysis of repeated survey data from the random samples. Early results in this area (see, for example, Jessen (1942), Patterson (1950), Rao and Graham (1964), Raj (1965) and Singh (1968)) have not assumed any relationship between successive values of the population mean. Blight and Scott (1973) and Scott and Smith (1974) have derived the estimates for the mean of a time dependent population using a first-order autoregressive model under the assumption that all the parameters of the model are known. Rodrigues and Bolfarine (1987) considered the prediction of the population

total (in a finite population) using a Bayesian approach based on the Kalman Filter estimating algorithm.

However, no results concerning the estimation of the parameters for the small areas under the dynamic superpopulation models wherein past information can be effectively used in conjunction with the current information, are available.

## 1.2 The Subject of This Dissertation

In this dissertation, we present a Bayesian analysis of the small domain data in repeated surveys. Estimation and prediction methodology have been developed for the infinite as well as finite population sampling situations. We consider simultaneous estimation of several strata means through a time series Bayesian estimation model.

In Chapter 2, we introduce a HB model for simultaneous estimation of several small areas means in infinite population sampling, assuming that the observations are available only at the aggregate level. The posterior distribution is provided through a fully Bayesian procedure which uses a Monte Carlo numerical integration technique, the Gibbs sampling algorithm. The resulting small area estimators “borrow strength” over time in addition to borrowing strength from other local areas. The methodology developed is applied to the problem of estimation of median income of four-person families by states.

In Chapter 3, we present a multivariate HB model for simultaneous estimation of several small areas means in an infinite population sampling, once again assuming that the observations are available only at the aggregate level. Our aim is to exploit the inter-relationship (if any) among the parameters in improving the small area estimators. We have also implemented the proposed analysis on the median income estimation problem considered in the previous chapter through the use of median incomes of three- and five-person families. The multivariate analysis in this case

shows definite improvement in the median income estimates when compared to the univariate analysis in Chapter 2.

Chapter 4 addresses the simultaneous estimation of several strata means at different time points, where each stratum at each time point contains a finite number of elements. We assume that the observations are available at the unit level. The HB model considered in this chapter can be regarded as an extension of the work of Datta and Ghosh (1991) who used general linear models for small area estimation. While the HB predictors of Datta and Ghosh (1991) utilize the data available at a given time, our HB predictors for small areas utilize the data available at a given time, along with all the data available at previous time points. Also, in addition to the current small area estimates, our model provides the retrospective estimates of the small areas at all the previous time points. This chapter also considers a special case of the time domain HB model for the finite population sampling, assuming known variance ratios. In such a set-up, the HB predictors have been shown to possess certain optimal properties. It has been shown that HB predictors are best unbiased predictors within the class of all unbiased predictors. Under milder assumptions, these predictors are best unbiased linear predictors. The stochastic dominance (cf. Hwang, 1985) of these predictors over the linear unbiased predictors for elliptically symmetric distributions have also been established in this chapter. Finally, we have also shown that these predictors are also best equivariant predictors for both the matrix loss as well as general quadratic loss, under suitable group of transformations for a class of elliptically symmetric distributions.



## CHAPTER 2

### ESTIMATION OF SMALL AREA MEANS IN TIME DOMAIN

#### 2.1 Introduction

This chapter considers the problem of small domain estimation based on data from repeated surveys. Estimation of the characteristic of interest, such as population mean, change in mean over two distinct time points, or other parameters of interest is considered by using a hierarchical Bayes (HB) approach. Such models, in contrast to the ones considered in Chapter 4, will be appropriate when observations are available at an aggregate level, rather than at a unit level. The previous work of Datta, Fay and Ghosh (1991) used a HB approach for analyzing such aggregate data for several small areas at a fixed point of time. Also, the empirical and subjective Bayes approaches considered earlier by Fay and Herriot (1979), Ericksen and Kadane (1985, 1987), Ericksen, Kadane and Tukey (1989) involved similar analysis of small domain data where no time series model was used. One common theme in all these papers was to “borrow strength” from similar other areas to obtain improved small area estimators. The present study borrows strength over time as well in addition to borrowing strength from other local areas.

On another front, Bayesian analysis of data from repeated surveys is also available in the literature which does not address the problem of simultaneous estimation from several small areas. Mention may be made in this context of the work of Blight and Scott (1972), Scott and Smith (1974) who considered estimation procedures from a

population whose mean varied over time, and applied standard normal time series methods. These results were later generalized by Rodrigues and Bolfarine (1987) who used a dynamic linear model, and used Kalman filter estimating algorithm. Our model encompasses all these models considered earlier as special cases when the number of local areas is one and the variance components, or at least the variance ratios are known.

In addition, the model that we consider in Section 2.2 has several novel features. Most of the models considered earlier for time series analysis of survey data assumed known variance components. The present work does not need such an assumption. Instead, the HB method assigns distributions (often non-informative) to these variance components, and thereby, in effect models the uncertainty due to unknown prior parameters. In consequence, the usual Kalman filter updating algorithm does not work any more, but the problem is addressed in a more realistic manner without invoking the assumption of known variances, or known ratios of variances. Also, the computational problems are overcome by using Monte-Carlo integration techniques such as Gibbs Sampling (see Gelfand and Smith (1990)).

Our model also bears some analogy to the one considered in Sallas and Harville (1988) when the number of local areas is one. However, while the Sallas – Harville approach is primarily an empirical Bayes (EB) approach, ours is a HB approach. The advantage of the present approach over an EB approach is that while the latter does not incorporate uncertainty due to estimation of the prior parameters, and as such typically underestimates the standard errors that need to be reported with the EB estimates, the former produces more honest estimates of the standard errors.

The outline of the remaining sections is as follows. Section 2 describes the basic hierarchical model, describes the Gibbs sampling algorithm, and computes the necessary posterior distributions needed to implement the Gibbs sampling procedure. This section also provides formulas for the posterior mean and variance via Gibbs



sampling. An alternative EB procedure is described in Section 3. The methodology developed in Section 2 is illustrated with the problem of estimation of median income of four-person families by states in Section 4.

## 2.2 Description of the Hierarchical Bayes Model

Let  $Y_{ij}$  denote the standard sample survey estimate of some characteristic  $\theta_{ij}$  such as the population mean based on a given sample at time  $j$  for the  $i^{th}$  small area ( $i = 1, \dots, m; j = 1, \dots, t$ ). Our objective is to optimally estimate a linear function  $\sum_{i=1}^m \sum_{j=1}^t l_{ij} \theta_{ij}$ , using all the data available upto time  $t$ . In particular, we may be interested in estimating current population means for all the small areas based on several sets of sample means available over time. To this end, we consider the following HB model:

- I.  $Y_{ij} | \theta_{ij} \stackrel{ind}{\sim} N(\theta_{ij}, V_{ij});$
- II.  $\theta_{ij} | \alpha, \mathbf{b}_j, \psi_j \stackrel{ind}{\sim} N(\mathbf{x}_{ij}^T \alpha + \mathbf{z}_{ij}^T \mathbf{b}_j, \psi_j) \ (i = 1, \dots, m; j = 1, \dots, t);$
- III.  $\mathbf{b}_j | \mathbf{b}_{j-1}, \mathbf{W} \stackrel{ind}{\sim} N(\mathbf{H}_j \mathbf{b}_{j-1}, \mathbf{W}) \ (j = 1, \dots, t);$
- IV. Marginally  $\alpha, \psi_1, \dots, \psi_t$ , and  $\mathbf{W}$  are mutually independent with
 
$$\alpha \sim Uniform(R^p),$$

$$R_j (= \psi_j^{-1}) \sim Gamma(\tfrac{1}{2}a_j, \tfrac{1}{2}h_j),$$

$$\mathbf{W} \sim inverseWishart(\mathbf{S}, k).$$

In (II),  $\mathbf{x}_{ij}$  and  $\mathbf{z}_{ij}$  are known design vectors of size  $p$  and  $q$  respectively.

We say that a random variable  $R$  has a *Gamma* ( $a, b$ ) distribution if it has a pdf of the form

$$f(r) \propto \exp(-\tfrac{1}{2}ar) r^{\frac{1}{2}b-1} I_{(0,\infty)}(r); a > 0, b > 0,$$

where  $I$  denotes the usual indicator function.

We say that a  $q \times q$ , positive definite random matrix  $\mathbf{W}$  has an *inverse Wishart* distribution with parameters  $S$  and  $k$  if it has a pdf of the form

$$f(\mathbf{w}) \propto |\mathbf{W}|^{-(q+k/2)} \exp \left[ -\frac{1}{2} \text{trace} \left( kS\mathbf{W}^{-1} \right) \right],$$

where  $k > 0$  is a known, scalar degrees of freedom parameter, and  $S$  a  $q \times q$  positive definite matrix. This means that  $\mathbf{W}^{-1}$  has a *Wishart* distribution with degrees of freedom  $k$  and scale parameter  $S$ .

In (IV), we shall allow the possibility of diffuse priors for  $R_j$  or  $\mathbf{W}$ , e.g.  $a_j = 0, b_j = 0$ , for  $(j = 1, \dots, t)$ ,  $S = 0$ .

In the case when  $t = 1$ , the suffix  $j$  can be omitted and  $Y_{ij} \equiv Y_i$  is the usual survey estimate of the small area parameter  $\theta_i$ . Write  $\psi_j = \psi$ ,  $\mathbf{x}_{ij} = \mathbf{x}_i$ . In such a case, we can write the model I-III as

$$\text{A. } Y_i | \theta_i \stackrel{\text{ind}}{\sim} N(\theta_i, V_i);$$

$$\text{B. } \theta_i | \mathbf{w} \stackrel{\text{ind}}{\sim} N(\mathbf{x}_i^T \boldsymbol{\alpha}, \psi + \mathbf{z}_i^T \mathbf{w} \mathbf{z}_i)$$

Special cases of this model are considered in Fay and Herriot (1979) and Ericksen and Kadane (1985, 1987) where  $\mathbf{z}_i = 0$ .

In the case when  $m = 1$ , the suffix  $i$  can be omitted and  $Y_{ij} \equiv Y_j$  is the usual survey estimate parameter  $\theta_j$  at time  $j$ . In such a case, our model in (I-III) can be compared with some of the models advocated for the analysis of repeated survey data. To see this, it is convenient to write (I-III) as:

$$\text{(i) } Y_j = \mathbf{x}_j^T \boldsymbol{\alpha} + b_j + e_j, \quad e_j \sim N(0, V_j),$$

$$\text{(ii) } b_j = H_j b_{j-1} + \omega_j, \quad \omega_j \sim N(0, W)$$

We first consider the time series model of Scott and Smith (1974) for analysis of repeated surveys. The special case of their model when a first-order autoregressive

process is assumed for the characteristic of interest  $\theta_j$  (which in this setup corresponds to  $b_j$ ) can be identified with (i) and (ii).

In general stages I and II of the model can be combined into a single stage

$$Y_{ij} = \mathbf{x}_j^T \boldsymbol{\alpha} + \mathbf{z}_{ij}^T \mathbf{b}_j + e_{ij}, \text{ where } e_{ij} \sim N(0, V_{ij} + \psi_j)$$

Our model bears resemblance to the one considered in Sallas and Harville (1988). However, there is a slight difference in the variance structure. To see this, combine stages I and II of the model. This leads to

$$Y_{ij} = \mathbf{x}_j^T \boldsymbol{\alpha} + \mathbf{z}_{ij}^T \mathbf{b}_j + e_{ij},$$

where  $e_{ij} \stackrel{\text{ind}}{\sim} N(0, V_{ij} + \psi_j)$ . In contrast, in Sallas and Harville (1988),  $V(\mathbf{e}_j) = \sigma^2 \mathbf{R}_j$ ,  $\mathbf{R}_j$  known,  $\sigma^2$  possibly unknown, where  $\mathbf{e}_j = (e_{1j}, \dots, e_{mj})^T$ .

We emphasize once again that the HB structure considered in this chapter assigns distributions to the parameters  $\boldsymbol{\alpha}, \psi_1, \dots, \psi_t$  and  $\mathbf{W}$ . To our knowledge, the problem has not been addressed to this level of generality before even within a Bayesian framework.

We now describe Gibbs sampling algorithm in its general framework. The Gibbs sampling is a Markovian updating scheme which requires sampling from full conditional distributions. Densities are denoted generally, by square brackets so that the joint, conditional and marginal densities appear as  $[U, V], [U|V], [V]$ , etc. Given a collection of random variables (real or vector valued)  $U_1, \dots, U_k$ , the joint density  $[U_1, \dots, U_k]$  is uniquely determined by  $[U_s|U_r, r \neq s], s = 1, \dots, k$ . under certain mild conditions (see Beseg 1974, pp 194–195). The interest is in finding the marginal distributions  $[U_s], s = 1, \dots, k$ .

Gibbs sampling algorithm is as follows: given an arbitrary set of starting values,  $U_1^{(0)}, \dots, U_k^{(0)}$ , draw  $U_1^{(1)}$  from  $[U_1|U_2^{(0)}, \dots, U_k^{(0)}]$ . Then draw  $U_2^{(1)}$  from  $[U_2|U_1^{(1)}, U_3^{(0)}, \dots, U_k^{(0)}], \dots, U_k^{(1)}$  from  $[U_k|U_1^{(1)}, \dots, U_{k-1}^{(1)}]$  to complete one iteration

of the scheme. After  $t$  such iterations, arrive at  $(U_1^{(t)}, \dots, U_k^{(t)})$ . There are two alternative ways to obtain i.i.d.  $k$ -tuples  $(U_{1g}^{(t)}, \dots, U_{kg}^{(t)}), (g = 1, \dots, G)$  from the joint distribution:

- Parallel: Replicate the entire process of  $t$  iterations in parallel  $G$  times storing only the last ( $t^{\text{th}}$ ) iterate to provide  $(U_{1g}^{(t)}, \dots, U_{kg}^{(t)}), g = 1, \dots, G$ .
- Sequential: Employ only one single long run of  $t \times G$  Gibbs iterates, storing every  $t^{\text{th}}$  iterate to provide  $(U_{1g}^{(t)}, \dots, U_{kg}^{(t)}), g = 1, \dots, G$ .

Gelfand and Smith (1990) adopted the parallel algorithm in simulating samples from posterior distributions. These simulated sample observations can then be used for estimation of any of the marginal densities. In particular, assuming that the conditional density  $f(U_s | U_r, r \neq s)$  is available in closed form, a Rao-Blackwellized sample-based estimate of the marginal density of  $U_s$  (see Gelfand and Smith (1991)) is given by

$$\hat{f}(U_s) = \frac{1}{G} \sum_{g=1}^G f(U_s | U_{rg}^{(t)}, r \neq s).$$

We will also be employing the parallel implementation of the Gibbs sampling algorithm in our analysis of the time domain HB model.

In order to implement the Gibbs sampling algorithm we require samples from the following complete conditional distributions:

- $\alpha | \mathbf{y}, \theta, \mathbf{b}_1, \dots, \mathbf{b}_t, r_1, \dots, r_t, \mathbf{W}$
- $\mathbf{b}_j | \mathbf{y}, \theta, \alpha, \mathbf{b}_1, \dots, \mathbf{b}_{j-1}, \mathbf{b}_{j+1}, \dots, \mathbf{b}_t, r_1, \dots, r_t, \mathbf{W}, j = 1, \dots, t$
- $r_1, \dots, r_t | \mathbf{y}, \theta, \mathbf{b}_1, \dots, \mathbf{b}_t, \mathbf{W}$



- $W|y, \theta, b_1, \dots, b_t, r_1, \dots, r_t$
- $\theta_{ij}|y, b_1, \dots, b_t, r_1, \dots, r_t, W, (i = 1, \dots, m; j = 1, \dots, t)$

The hierarchical structure of the model implies that

- (i)  $\alpha|y, \theta, b_1, \dots, b_t, r_1, \dots, r_t, W \sim \alpha|\theta, b_1, \dots, b_t, r_1, \dots, r_t$
- (iia)  $b_j|y, \theta, \alpha, b_{j' \neq j}, r_1, \dots, r_t, W \sim b_j|\theta, \alpha, b_{j-1}, b_{j+1}, r_j, W$   
( $j = 1, \dots, t-1$ )
- (iib)  $b_t|y, \theta, \alpha, b_{j' \neq j}, r_1, \dots, r_t, W \sim b_t|\theta, \alpha, b_{t-1}, r_t, W$
- (iii)  $r_j|y, \theta, \alpha, b_1, \dots, b_t, W \sim r_j|\alpha, \theta, b_1, \dots, b_t (j = 1, \dots, t)$
- (iv)  $W|y, \theta, b_1, \dots, b_t, r_1, \dots, r_t \sim W|b_1, \dots, b_t$
- (v)  $\theta_{ij}|y, b_1, \dots, b_t, r_1, \dots, r_t, W \sim \theta_{ij}|y_{ij}, r_j, b_j, \alpha$   
( $i = 1, \dots, m; j = 1, \dots, t$ )

Formal derivation of the complete conditional distributions is now straightforward and is omitted. The complete conditional distributions are given by

$$(i) \quad \alpha|y, \theta, b_1, \dots, b_t, r_1, \dots, r_t, W \\ \sim N_p[(\sum_{i=1}^m \sum_{j=1}^t r_j \mathbf{x}_{ij}^T \mathbf{x}_{ij})^{-1} \sum_{i=1}^m \sum_{j=1}^t r_j (\theta_{ij} - \mathbf{z}_{ij}^T \mathbf{b}_j) \mathbf{x}_{ij}, \\ (\sum_{i=1}^m \sum_{j=1}^t r_j \mathbf{x}_{ij}^T \mathbf{x}_{ij})^{-1}]$$

(ii a) For ( $j = 1, \dots, t-1$ ) :

$$b_j|y, \theta, \alpha, b_1, \dots, b_{j-1}, b_{j+1}, \dots, b_t, r_1, \dots, r_t, W \\ \sim N_q[r_j(\sum_{i=1}^m \mathbf{z}_{ij}^T \mathbf{z}_{ij} + \mathbf{H}_{j+1}^T \mathbf{W}^{-1} \mathbf{H}_{j+1} + \mathbf{W}^{-1})^{-1} \\ \times \{(\sum_{i=1}^m \mathbf{z}_{ij}^T \mathbf{z}_{ij})^{-1} \sum_{i=1}^m \mathbf{z}_{ij}^T (\theta_{ij} - \mathbf{x}_{ij}^T \alpha) \\ + \mathbf{H}_{j+1}^T \mathbf{W}^{-1} \mathbf{b}_{j+1} + \mathbf{W}^{-1} \mathbf{H}_j \mathbf{b}_{j-1}\}, \\ (\sum_{i=1}^m r_j \mathbf{z}_{ij}^T \mathbf{z}_{ij} + \mathbf{H}_{j+1}^T \mathbf{W}^{-1} \mathbf{H}_{j+1} + \mathbf{W}^{-1})^{-1}]$$

$$\begin{aligned}
(ii) \quad & b_t | \mathbf{y}, \boldsymbol{\theta}, \boldsymbol{\alpha}, b_1, \dots, b_t, r_1, \dots, r_t, \mathbf{W} \\
& \sim N_q [r_t \sum_{i=1}^m \mathbf{z}_{it}^T \mathbf{z}_{it} + \mathbf{W}^{-1}]^{-1} \\
& \times \{ (\sum_{i=1}^m \mathbf{z}_{it}^T \mathbf{z}_{it})^{-1} \sum_{i=1}^m \mathbf{z}_{it}^T (\theta_{it} - \mathbf{x}_{it}^T \boldsymbol{\alpha}) \\
& + \mathbf{W}^{-1} \mathbf{H}_t \mathbf{b}_{t-1} \}, (\sum_{i=1}^m r_j \mathbf{z}_{it}^T \mathbf{z}_{it} + \mathbf{W}^{-1})^{-1}
\end{aligned}$$

$$\begin{aligned}
(iii) \quad & r_1, \dots, r_t | \mathbf{y}, \boldsymbol{\theta}, b_1, \dots, b_t, \mathbf{W} \\
& \stackrel{ind}{\sim} \text{Gamma}(\frac{1}{2}(a_j + \sum_{i=1}^m \sum_{j=1}^t \theta_{ij} - \mathbf{x}_{ij}^T \boldsymbol{\alpha} - \mathbf{z}_{ij}^T \mathbf{b}_j)^2, \frac{1}{2}(h_j + m))
\end{aligned}$$

$$\begin{aligned}
(iv) \quad & \mathbf{W} | \mathbf{y}, \boldsymbol{\theta}, b_1, \dots, b_t, r_1, \dots, r_t \\
& \sim \text{Inverse Wishart}(\mathbf{S} + \sum_{j=1}^t (\mathbf{b}_j - \mathbf{H}_j \mathbf{b}_{j-1})(\mathbf{b}_j - \mathbf{H}_j \mathbf{b}_{j-1})^T, k + t)
\end{aligned}$$

$$\begin{aligned}
(v) \quad & \theta_{ij} | \mathbf{y}, b_1, \dots, b_t, r_1, \dots, r_t, \mathbf{W} \\
& \sim N[(V_{ij}^{-1} + r_j)^{-1} (V_{ij}^{-1} y_{ij} + r_j (\mathbf{x}_{ij}^T \boldsymbol{\alpha} + \mathbf{z}_{ij}^T \mathbf{b}_j), (V_{ij}^{-1} + r_j)^{-1}]
\end{aligned}$$

Our objective is to find the posterior expectation and variance of  $\theta_{ij}$  given  $\mathbf{y}$ . A Rao-Blackwellized estimate for  $E(\theta_{ij} | \mathbf{y})$  based on simulated samples from the above conditional distributions can be approximated by

$$\frac{1}{G} \sum_{g=1}^G E(\theta_{ij} | \mathbf{y}, \boldsymbol{\alpha} = \boldsymbol{\alpha}_g^{(t)}, \mathbf{W} = \mathbf{W}_g^{(t)}, \mathbf{b}_j = \mathbf{b}_{jg}, r_j = r_{jg}, j = 1, \dots, t)$$

The posterior variance can be approximated by

$$\begin{aligned}
& \frac{1}{G-1} \left[ E^2(\theta_{ij} | \mathbf{y}, \boldsymbol{\alpha} = \boldsymbol{\alpha}_g^{(t)}, \mathbf{W} = \mathbf{W}_g^{(t)}, \mathbf{b}_j = \mathbf{b}_{jg}, r_j = r_{jg}, j = 1, \dots, t) \right. \\
& \left. - \frac{1}{G} \left( \sum_{g=1}^G E(\theta_{ij} | \mathbf{y}, \boldsymbol{\alpha} = \boldsymbol{\alpha}_g^{(t)}, \mathbf{W} = \mathbf{W}_g^{(t)}, \mathbf{b}_j = \mathbf{b}_{jg}, r_j = r_{jg}, j = 1, \dots, t) \right)^2 \right] \\
& + \frac{1}{G} V(\theta_{ij} | \mathbf{y}, \boldsymbol{\alpha} = \boldsymbol{\alpha}_g^{(t)}, \mathbf{W} = \mathbf{W}_g^{(t)}, \mathbf{b}_j = \mathbf{b}_{jg}, r_j = r_{jg}, j = 1, \dots, t)
\end{aligned}$$

### 2.3 Empirical Bayes Estimation

In an EB formulation, we assume (I-III) of the model, but not (IV). Also, it is assumed that  $\mathbf{W}$  is a known matrix. In such a case, we may write the model as

$$\begin{aligned} Y_{ij} | \theta_{ij} &\stackrel{\text{ind}}{\sim} N(\theta_{ij}, V_{ij}) \\ \theta_{ij} &\stackrel{\text{ind}}{\sim} N \left( \mathbf{x}_{ij}^T \boldsymbol{\alpha}, \psi_j + \mathbf{z}_{ij}^T (\mathbf{W} + \sum_{l=1}^{j-1} \mathbf{H}_l \mathbf{H}_l^T) \mathbf{z}_{ij} \right) \end{aligned} \quad (2.3.1)$$

assuming known  $\mathbf{b}_0$ , i.e.,  $\mathbf{b}_0 = \mathbf{0}$  without any loss of generality.

It follows from 2.3.1 that

$$\begin{bmatrix} Y_{ij} \\ \theta_{ij} \end{bmatrix} \sim N \left[ \begin{pmatrix} \mathbf{x}_{ij}^T \boldsymbol{\alpha} \\ \mathbf{x}_{ij}^T \boldsymbol{\alpha} \end{pmatrix}, \begin{pmatrix} D_{ij} + V_{ij} & D_{ij} \\ D_{ij} & D_{ij} \end{pmatrix} \right] \quad (2.3.2)$$

where

$$D_{ij} = \psi_j + \mathbf{z}_{ij}^T (\mathbf{W} + \sum_{l=1}^{j-1} \mathbf{H}_l \mathbf{H}_l^T) \mathbf{z}_{ij} \quad (2.3.3)$$

From (2.3.2), the posterior distribution of  $\theta_{ij}$  given  $Y_{ij}$  can be obtained as

$$\theta_{ij} | Y_{ij} \stackrel{\text{ind}}{\sim} N \left[ \mathbf{x}_{ij}^T \boldsymbol{\alpha} + D_{ij} (D_{ij} + V_{ij})^{-1} (Y_{ij} - \mathbf{x}_{ij}^T \boldsymbol{\alpha}), D_{ij} V_{ij} (D_{ij} + V_{ij})^{-1} \right] \quad (2.3.4)$$

The Bayes estimator of  $\theta_{ij}$  under any quadratic loss is its posterior mean, and is given by

$$\hat{\theta}_{ij}^B = \mathbf{x}_{ij}^T \boldsymbol{\alpha} + D_{ij} (D_{ij} + V_{ij})^{-1} (Y_{ij} - \mathbf{x}_{ij}^T \boldsymbol{\alpha}) \quad (2.3.5)$$

which may also be seen as a weighted average of the survey estimate  $Y_{ij}$  and  $\mathbf{x}_{ij}^T \boldsymbol{\alpha}$ .

In an EB analysis, at least one of the parameters  $\boldsymbol{\alpha}$  and  $\psi_j$ , ( $j = 1, \dots, t$ ); is unknown and needs to be estimated from the marginal distribution of the  $Y_{ij}$ 's.

First we estimate  $\boldsymbol{\alpha}$ . Marginally

$$Y_{ij} \stackrel{ind}{\sim} N(\mathbf{x}_{ij}^T \boldsymbol{\alpha}, D_{ij} + V_{ij})$$

Using the method of maximum likelihood,  $\boldsymbol{\alpha}$  can be estimated by

$$\hat{\boldsymbol{\alpha}} = \left[ \sum_{i=1}^m \sum_{j=1}^t \mathbf{x}_{ij} \mathbf{x}_{ij}^T / (D_{ij} + V_{ij}) \right]^{-1} \left[ \sum_{i=1}^m \sum_{j=1}^t \mathbf{x}_{ij}^T Y_{ij} / (D_{ij} + V_{ij}) \right] \quad (2.3.6)$$

An EB estimator of  $\theta_{ij}$  is obtained by substituting  $\hat{\boldsymbol{\alpha}}$  for  $\boldsymbol{\alpha}$  in (2.3.5) and is given by

$$\hat{\theta}_{ij}^{EB} = \mathbf{x}_{ij}^T \hat{\boldsymbol{\alpha}} + D_{ij} (D_{ij} + V_{ij})^{-1} (Y_{ij} - \mathbf{x}_{ij}^T \hat{\boldsymbol{\alpha}}) \quad (2.3.7)$$

The EB approach now attempts to estimate  $\psi_j$  by maximum likelihood method or some other standard method. For example, noting that

$$E \left[ \sum_{i=1}^m (Y_{ij} - \mathbf{x}_{ij}^T \tilde{\boldsymbol{\alpha}})^2 / (D_{ij} + V_{ij}) \right] = m - q,$$

one can solve iteratively

$$\sum_{i=1}^m (Y_{ij} - \mathbf{x}_{ij}^T \tilde{\boldsymbol{\alpha}})^2 / (D_{ij} + V_{ij}) = m - q \quad (2.3.8)$$

in conjunction with (2.3.6), to find  $\hat{\psi}_j$  if a positive solution for  $\hat{\psi}_j$  is found and letting  $\hat{\psi}_j = 0$ , when no positive solution could be found.

We observe that the EB approach and the HB approach (considered in Section 2.2) will lead to the same point estimator of  $\theta_{ij}$ . However, the EB procedure by ignoring the uncertainty in estimating the unknown parameters will lead to the underestimation of the true posterior standard errors. The complete HB analysis though involved, provide more reliable estimates of standard errors. It is for this reason, we will be employing only the HB approach in the illustrative example considered in Section 2.4.



## 2.4 An Application of Hierarchical Bayes Analysis

In this section, we apply the HB procedure developed in Section 2.2. to an actual data set. Our application concerns estimation of median income for four-person families by state. Before employing the HB model for median income estimation, we will briefly give a background of this problem. The details are given in Fay (1987).

The U.S. Department of Health and Human Services (HHS) administers a program of energy assistance to low-income families. Eligibility for the program is determined by a formula where the most important variable is an estimate of the current median income for four-person families by states (the 50 states and the District of Columbia).

The Bureau of the Census, by an informal agreement, has provided estimates of the median income for the four-person families by states to HHS through a linear regression methodology since the latter part of the 1970s. Two basic approaches have been used with change in methodology occurring for income year 1985. In this earlier method, sample estimates  $Y$  of the state medians for the most current year  $c$ , as well as the associated standard errors are first obtained through Current Population Survey (CPS). The estimates  $Y$  were used as dependent variables in a linear regression procedure with the single predictor variable

$$\begin{aligned} \text{Adjusted census median}(c) = & [\text{BEA PCI}(c)/\text{BEA PCI}(b)] \\ & \times \text{Census median}(b), \end{aligned} \tag{2.4.9}$$

where census median ( $b$ ) represents the median income of the four-person families in the state for the base year  $b$  from the most recently available decennial census. In the above BEA PCI( $c$ ) and BEA PCI( $b$ ) represent respectively estimates of per-capita income produced by the Bureau of Economic Analysis of the U.S. Department of Commerce for the current year  $c$ , and the base-year  $b$ , respectively. Thus (2.4.9) attempts to adjust the base year census median by the proportional growth in the BEA PCI to arrive at the current year adjusted median. Moreover, in developing the

model, the states were divided on the basis of the population into four groups of 12 or 13 states each. Model variances for each were computed as the average of squared residuals of 1969 census medians fitted by a linear regression with an intercept term along with the adjusted census median for 1969 (as given in (2.4.9)) as a covariate.

Next a weighted average of the CPS sample estimate of the current median income and the regression estimates is obtained, weighting the regression estimate inversely proportional to the model error due to the lack of fit for the census values of 1969 by the model with 1959 used as the base year, and weighting the sample estimate inversely proportional to the sampling variance. From a Bayesian angle, the above composite estimate is obtainable from the simple hierarchical model of Lindley and Smith (1972). Finally, the above composite estimate was constrained not to deviate by more than one standard deviation away from the corresponding sample estimates. Thus the final estimates were analogous to the "Limited Translation Estimates" as discussed for example in Efron and Morris (1971) or Fay and Herriot (1979).

The model was reviewed once the median income of four-families by state in 1979 became available from the 1980 census. The following modifications to the previous modelling were suggested on the basis of this review.

First, it was decided to consider a single level of model error across all the states. The reason was that the differences between the estimated model errors for the larger and smaller states greatly exaggerated more current differences. Second, the predictive value of the regression appeared usefully improved if in addition to the adjusted census median ( $c$ ), the census median ( $b$ ) was also included as a second independent variable. The idea was to adjust for any possible overstatement of the effect of change in BEA income upon the median incomes of four-person families. Also, the one standard deviation constraint was removed. This matter is discussed in Fay (1987).

So far, the entire time series of the sample median income estimates available for each year from the preceding census have not been used to obtain the median income estimates for the four-person families. Our model in Section 2.2 exploits the past time series information available at a given time in addition to the recent most census medians. Before we describe the hierarchical model for median income estimation, we need to introduce a few notations. Let  $Y_{ij}$  denote the median income of four-person families in state  $i$ . The true median corresponding to  $Y_{ij}$  is denoted by  $\theta_{ij}$ . Let  $x_{ij1}$  and  $x_{ij2}$  denote respectively the adjusted census median income for the year  $j$  (as given in (2.4.9)), and the base-year census median income for four-person families in the  $i^{th}$  state ( $i = 1, \dots, 51$ ). Finally, let

$$\mathbf{x}_{ij}^T = (1 \ x_{ij1} \ x_{ij2}), \ (i = 1, \dots, 51; \ j = 1, \dots, 10), \quad (2.4.10)$$

and denote by  $\boldsymbol{\alpha} = (\alpha_1 \ \alpha_2 \ \alpha_3)^T$  the vector of regression coefficients. The following hierarchical model is used to estimate the  $\theta_{ij}$ 's:

I.  $Y_{ij} | \theta_{ij}, v_{ij}$  are mutually independent with

$$Y_{ij} \stackrel{\text{ind}}{\sim} N(\theta_{ij}, v_{ij}), \ (i = 1, \dots, 51; \ j = 1, \dots, 10)$$

II.  $\theta_{ij} | \boldsymbol{\alpha}, b_j, \psi_j \stackrel{\text{ind}}{\sim} N(\mathbf{x}_{ij}^T \boldsymbol{\alpha} + b_j, \psi_j), \ (i = 1, \dots, 51; \ j = 1, \dots, 10)$

III.  $b_j | b_{j-1}, \tau^2 \stackrel{\text{ind}}{\sim} N(b_{j-1}, \tau^2) \ (j = 1, \dots, 10)$

IV. Marginally  $\boldsymbol{\alpha}, \psi_1, \dots, \psi_{10}$ , and  $\tau^2$  are mutually independent with

$$\boldsymbol{\alpha} \sim \text{uniform}(R^3),$$

$$r_j = \psi_j^{-1} \sim \text{Gamma}(0, -1), \ (j = 1, \dots, 10)$$

$$r_0 = (\tau^2)^{-1} \sim \text{Gamma}(0, 0),$$

To find the estimates of the median income of four-person families and their standard errors, we need to use the Gibbs sampling algorithm and require samples from the

following complete conditional distributions:

- $\alpha|\mathbf{y}, \boldsymbol{\theta}, b_1, \dots, b_{10}, r_1, \dots, r_{10}, r_0$
- $b_j|\mathbf{y}, \boldsymbol{\theta}, \alpha, b_1, \dots, b_{j-1}, b_{j+1}, \dots, b_{10}, r_1, \dots, r_{10}, r_0, j = 1, \dots, 10$
- $r_1, \dots, r_{10}|\mathbf{y}, \boldsymbol{\theta}, b_1, \dots, b_{10},$
- $r_0|\mathbf{y}, \boldsymbol{\theta}, b_1, \dots, b_{10}, r_1, \dots, r_{10}$
- $\theta_{ij}|\mathbf{y}, b_1, \dots, b_{10}, r_1, \dots, r_{10}, r_0, (i = 1, \dots, 51; j = 1, \dots, 10)$

Again, the hierarchical structure of the model implies that

- (i)  $\alpha|\mathbf{y}, \boldsymbol{\theta}, b_1, \dots, b_{10}, r_1, \dots, r_{10}, r_0 \sim \alpha|\boldsymbol{\theta}, b_1, \dots, b_{10}, r_1, \dots, r_{10}$
- (iia)  $b_j|\mathbf{y}, \boldsymbol{\theta}, \alpha, b_{j-1}, b_{j+1}, r_j, r_0 \sim b_j|\boldsymbol{\theta}, \alpha, b_{j-1}, b_{j+1}, r_j, r_0$   
( $j = 1, \dots, 9$ )
- (iib)  $b_{10}|\mathbf{y}, \boldsymbol{\theta}, \alpha, b_1, \dots, b_9, r_1, \dots, r_{10}, r_0 \sim b_{10}|\boldsymbol{\theta}, \alpha, b_9, r_{10}, r_0$
- (iii)  $r_j|\mathbf{y}, \boldsymbol{\theta}, \alpha, b_1, \dots, b_{10}, r_0 \sim r_j|\alpha, \boldsymbol{\theta}, b_1, \dots, b_{10} (j = 1, \dots, 10)$
- (iv)  $r_0|\mathbf{y}, \boldsymbol{\theta}, b_1, \dots, b_{10}, r_1, \dots, r_{10} \sim r_0|b_1, \dots, b_{10}$
- (v)  $\theta_{ij}|\mathbf{y}, b_1, \dots, b_{10}, r_1, \dots, r_{10}, r_0 \sim \theta_{ij}|y_{ij}, r_j, b_j, \alpha$   
( $i = 1, \dots, m; j = 1, \dots, 10$ )

These distributions under conjugate priors follow from standard normal and gamma distribution theory and are as follows:

$$\begin{aligned}
 (i) \quad & \alpha|\mathbf{y}, \boldsymbol{\theta}, b_1, \dots, b_{10}, r_1, \dots, r_{10}, r_0 \\
 & \sim N_3[(\sum_{i=1}^{51} \sum_{j=1}^{10} r_j \mathbf{x}_{ij}^T \mathbf{x}_{ij})^{-1} \sum_{i=1}^{51} \sum_{j=1}^{10} r_j \mathbf{x}_{ij}^T (\theta_{ij} - \mathbf{b}_j), \\
 & (\sum_{i=1}^{51} \sum_{j=1}^{10} r_j \mathbf{x}_{ij}^T \mathbf{x}_{ij})^{-1}]
 \end{aligned}$$



(ii a) For  $(j = 1, \dots, 9)$  :

$$b_j | \mathbf{y}, \boldsymbol{\theta}, \boldsymbol{\alpha}, b_{j-1}, b_{j+1}, r_j, r_0 \\ \sim N[(51 r_j + 2 r_0)^{-1} \{r_j \sum_{i=1}^{51} (\theta_{ij} - \mathbf{x}_{ij}^T \boldsymbol{\alpha}) \\ + r_0(b_{j+1} + b_{j-1})\}, (51 r_j + r_0)^{-1}]$$

(ii b)  $b_{10} | \mathbf{y}, \boldsymbol{\theta}, \boldsymbol{\alpha}, b_9, r_{10}, r_0$

$$\sim N[(51 r_{10} + r_0)^{-1} \{r_{10} \sum_{i=1}^{51} (\theta_{i10} - \mathbf{x}_{i,10}^T \boldsymbol{\alpha}) + r_0 b_9\}, \\ (51 r_{10} + r_0)^{-1}]$$

(iii)  $r_1, \dots, r_{10} | \mathbf{y}, \boldsymbol{\theta}, b_1, \dots, b_{10}, r_0$

$$\stackrel{ind}{\sim} \text{Gamma}(\frac{1}{2} \sum_{i=1}^{51} \sum_{j=1}^{10} (\theta_{ij} - \mathbf{x}_{ij}^T \boldsymbol{\alpha} - b_j)^2, 25)$$

(iv)  $r_0 | \mathbf{y}, \boldsymbol{\theta}, b_1, \dots, b_{10}, r_1, \dots, r_{10}$

$$\sim \text{Gamma}(\frac{1}{2} \sum_{j=1}^{10} (b_j - b_{j-1})(b_j - b_{j-1})^T, 5)$$

(v)  $\theta_{ij} | \mathbf{y}, b_1, \dots, b_{10}, r_1, \dots, r_{10}, r_0$

$$\sim N[(v_{ij}^{-1} + r_j)^{-1} (v_{ij}^{-1} y_{ij} + r_j (\mathbf{x}_{ij}^T \boldsymbol{\alpha} + b_j)), \\ (v_{ij}^{-1} + r_j)^{-1}]$$

Simulations from these complete conditional distributions can be made using the standard random number generators from the normal and gamma distributions. We can now estimate the marginal posterior density of  $\boldsymbol{\theta}_{ij} | \mathbf{y}$  as described earlier in Section 2.2. We find estimates of the four-person families for each year from 1979–1989 (except for 1980) using all the sample estimates available since 1979 census by running 50 iterations of 100–2000 replications each. The necessary convergence is achieved after 500 samples. We denote these HB estimates by HB<sup>1</sup>. As a special case of our model, we have also computed the median income estimates for 1989 utilizing only the census median income figures, the CPS median income estimates for 1989, and



the PCI's for the years 1979 and 1989. We denote these HB estimates by HB<sup>2</sup>. The HB<sup>2</sup> estimates are not based on any time series model, and are based on a univariate version of the HB model considered in Datta, Fay and Ghosh (1991). Table 2.1 provides the statewise sample median incomes as well as the HB<sup>1</sup> and HB<sup>2</sup> estimates for 1989. The standard errors associated with these estimates are given in Table 2.2. Also, for most states, the time domain HB analysis provides reduced estimates of standard errors, when compared to the corresponding standard errors under the non time-series model of Datta, Fay and Ghosh (1991). The time domain HB procedure results in substantial reduction of the standard errors when compared to the sample estimates.

The median income estimates along with their standard errors for the year 1979 are provided in Table 2.3. Since in addition to the sample estimates, the 1979 incomes are also available from the 1980 census data, we compare the retrospective estimates obtained for 1979 by using the time series model against the 1980 census, treating the latter as the "truth." The 1980 Census figures are labelled as Truth in Table-2.3.

Suppose  $e_{iTR}$  denotes the true median income for the  $i^{th}$  state, and  $e_i$  any estimate of  $e_{iTR}$ . Then for the estimate  $e = (e_1, \dots, e_{51})^T$  of  $e_{TR} = (e_{1TR}, \dots, e_{51TR})^T$ , let

$$\text{Average Relative Bias} = (51)^{-1} \sum_{i=1}^{51} |e_i - e_{iTR}| / e_{iTR}$$

$$\text{Average Squared Relative Bias} = (51)^{-1} \sum_{i=1}^{51} |e_i - e_{iTR}|^2 / e_{iTR}^2$$

$$\text{Average Absolute Bias} = (51)^{-1} \sum_{i=1}^{51} |e_i - e_{iTR}|$$

$$\text{Average Squared Deviation} = (51)^{-1} \sum_{i=1}^{51} (e_i - e_{iTR})^2$$

The above quantities are provided in Table 2.4 for the sample estimates and HB<sup>1</sup> estimates. The HB<sup>1</sup> estimates are clearly superior to the sample median estimates for each state. In terms of the average relative bias, the HB<sup>1</sup> estimates improve on the sample medians by 80.32 %. Using the criterion of average squared relative bias, the corresponding percentage improvement is 97.01. Similarly, for average absolute bias

and average squared deviations as the criteria, the respective percentage improvement of the HB<sup>1</sup> estimates over the sample medians are 80.88 and 96.27. Thus time series HB procedure improves tremendously on the sample estimates.

Table 2.1. Median Income for 4-Person Family by State in 1989  
(In Dollars )

| State | Sample<br>Median | HB <sup>1</sup> | HB <sup>2</sup> |
|-------|------------------|-----------------|-----------------|
| 1     | 37700            | 37332           | 37987           |
| 2     | 52720            | 47829           | 48232           |
| 3     | 41489            | 39604           | 40086           |
| 4     | 55729            | 52661           | 52817           |
| 5     | 47106            | 43424           | 43629           |
| 6     | 55048            | 53203           | 53020           |
| 7     | 43434            | 43855           | 43879           |
| 8     | 53128            | 53436           | 53305           |
| 9     | 40990            | 40745           | 40762           |
| 10    | 44241            | 42589           | 42550           |
| 11    | 40297            | 39736           | 39604           |
| 12    | 42590            | 43075           | 42817           |
| 13    | 43260            | 43101           | 42907           |
| 14    | 42226            | 41111           | 41006           |
| 15    | 45432            | 43631           | 43439           |
| 16    | 38711            | 37773           | 37713           |
| 17    | 39232            | 38932           | 38976           |
| 18    | 36966            | 34946           | 35068           |
| 19    | 32610            | 32116           | 32387           |
| 20    | 40894            | 38326           | 38419           |
| 21    | 42781            | 38820           | 38821           |
| 22    | 39819            | 41924           | 41724           |
| 23    | 48252            | 50346           | 49893           |
| 24    | 36587            | 40270           | 40384           |
| 25    | 49718            | 45511           | 45656           |

Table 2.1. continued

| State | Sample<br>Median | HB <sup>1</sup> | HB <sup>2</sup> |
|-------|------------------|-----------------|-----------------|
| 26    | 31604            | 32141           | 32217           |
| 27    | 37423            | 37399           | 37564           |
| 28    | 31834            | 34407           | 34626           |
| 29    | 45876            | 41050           | 41479           |
| 30    | 37194            | 37827           | 37876           |
| 31    | 32186            | 33546           | 33779           |
| 32    | 31327            | 33843           | 34041           |
| 33    | 32245            | 34079           | 34272           |
| 34    | 31330            | 31559           | 31922           |
| 35    | 33317            | 32480           | 32918           |
| 36    | 36219            | 35123           | 35118           |
| 37    | 40358            | 35680           | 35865           |
| 38    | 32679            | 34247           | 34117           |
| 39    | 29586            | 32598           | 32602           |
| 40    | 32227            | 33030           | 33161           |
| 41    | 38931            | 36451           | 35852           |
| 42    | 36732            | 39591           | 39294           |
| 43    | 29761            | 31238           | 31462           |
| 44    | 38969            | 38828           | 38868           |
| 45    | 37938            | 36913           | 36942           |
| 46    | 34579            | 39129           | 38722           |
| 47    | 42392            | 41821           | 41559           |
| 48    | 41419            | 39342           | 39256           |
| 49    | 41121            | 42010           | 41820           |
| 50    | 41379            | 48607           | 47146           |
| 51    | 44999            | 44785           | 44369           |



Table 2.2. Estimated Standard Errors of Median Income Estimates  
for 4-Person Family by State in 1989

(In Dollars )

| State | Sample<br>Median | HB <sup>1</sup> | HB <sup>2</sup> |
|-------|------------------|-----------------|-----------------|
| 1     | 3230             | 2010            | 2147            |
| 2     | 4132             | 2327            | 2450            |
| 3     | 2848             | 1936            | 2011            |
| 4     | 2452             | 1946            | 1999            |
| 5     | 4512             | 2289            | 2355            |
| 6     | 3893             | 2245            | 2379            |
| 7     | 1512             | 1302            | 1318            |
| 8     | 1933             | 1575            | 1635            |
| 9     | 1584             | 1335            | 1349            |
| 10    | 1845             | 1562            | 1583            |
| 11    | 2970             | 1931            | 1974            |
| 12    | 1763             | 1440            | 1472            |
| 13    | 1597             | 1341            | 1374            |
| 14    | 2310             | 1722            | 1757            |
| 15    | 3251             | 2030            | 2083            |
| 16    | 2258             | 1698            | 1734            |
| 17    | 3677             | 2098            | 2141            |
| 18    | 2249             | 1757            | 1794            |
| 19    | 2130             | 1630            | 1679            |
| 20    | 2502             | 1869            | 1899            |
| 21    | 3359             | 2134            | 2184            |
| 22    | 3699             | 2117            | 2164            |
| 23    | 4665             | 2314            | 2450            |
| 24    | 4645             | 2280            | 2336            |
| 25    | 4349             | 2292            | 2362            |

Table 2.2. Continued

| State | Sample<br>Median | HB <sup>1</sup> | HB <sup>2</sup> |
|-------|------------------|-----------------|-----------------|
| 26    | 2278             | 1695            | 1736            |
| 27    | 1513             | 1289            | 1311            |
| 28    | 2965             | 1971            | 2020            |
| 29    | 4385             | 2295            | 2377            |
| 30    | 1397             | 1232            | 1242            |
| 31    | 2885             | 1915            | 1968            |
| 32    | 2446             | 1830            | 1870            |
| 33    | 2690             | 1869            | 1913            |
| 34    | 2392             | 1733            | 1804            |
| 35    | 2435             | 1762            | 1836            |
| 36    | 3314             | 2039            | 2099            |
| 37    | 3923             | 2244            | 2311            |
| 38    | 1552             | 1405            | 1394            |
| 39    | 2557             | 1896            | 1918            |
| 40    | 2268             | 1691            | 1727            |
| 41    | 3516             | 2158            | 2356            |
| 42    | 3261             | 2045            | 2088            |
| 43    | 2811             | 1902            | 1962            |
| 44    | 5003             | 2293            | 2346            |
| 45    | 3209             | 2004            | 2050            |
| 46    | 2863             | 2059            | 2084            |
| 47    | 2525             | 1788            | 1842            |
| 48    | 3056             | 1993            | 2040            |
| 49    | 1586             | 1359            | 1369            |
| 50    | 3637             | 2299            | 2665            |
| 51    | 4805             | 2278            | 2369            |

Table 2.3. Median Income Estimates for 4-Person Family by State in 1979

| State | Truth | (In Dollars )    |      | HB <sup>1</sup> |     |
|-------|-------|------------------|------|-----------------|-----|
|       |       | Sample<br>Median | SE   | Median          | SE  |
| 1     | 18319 | 18084            | 1533 | 17967           | 516 |
| 2     | 22027 | 23129            | 1647 | 21919           | 495 |
| 3     | 19424 | 18438            | 1642 | 19055           | 508 |
| 4     | 23772 | 25324            | 1414 | 23805           | 501 |
| 5     | 22107 | 23597            | 2409 | 21967           | 511 |
| 6     | 25712 | 25037            | 1431 | 25596           | 509 |
| 7     | 22669 | 21852            | 766  | 22329           | 423 |
| 8     | 26014 | 25486            | 1264 | 25913           | 502 |
| 9     | 22266 | 23326            | 878  | 22330           | 460 |
| 10    | 23279 | 22397            | 843  | 22963           | 439 |
| 11    | 23014 | 22716            | 1041 | 22826           | 451 |
| 12    | 25410 | 24294            | 1025 | 25168           | 486 |
| 13    | 25111 | 24103            | 1006 | 24872           | 477 |
| 14    | 23320 | 23270            | 1202 | 23179           | 467 |
| 15    | 24044 | 25137            | 1571 | 24021           | 501 |
| 16    | 22351 | 23851            | 1592 | 22294           | 499 |
| 17    | 21891 | 20743            | 1255 | 21563           | 478 |
| 18    | 20511 | 18567            | 1535 | 20094           | 506 |
| 19    | 18674 | 19826            | 1588 | 18449           | 522 |
| 20    | 21438 | 22603            | 1802 | 21297           | 501 |
| 21    | 22127 | 22014            | 1586 | 21930           | 487 |
| 22    | 23627 | 21860            | 1900 | 23388           | 510 |
| 23    | 26203 | 26235            | 1722 | 26173           | 528 |
| 24    | 21862 | 20232            | 3801 | 21623           | 520 |
| 25    | 22757 | 24160            | 1418 | 22736           | 494 |

Table 2.3. continued

| State | Truth | Sample |      | HB <sup>1</sup> |     |
|-------|-------|--------|------|-----------------|-----|
|       |       | Median | SE   | Median          | SE  |
| 26    | 20214 | 18274  | 1380 | 19758           | 506 |
| 27    | 19772 | 20296  | 1012 | 19611           | 465 |
| 28    | 19944 | 19282  | 1795 | 19624           | 506 |
| 29    | 20668 | 22687  | 1196 | 20703           | 517 |
| 30    | 21086 | 19675  | 1042 | 20650           | 474 |
| 31    | 19685 | 18657  | 1285 | 19295           | 489 |
| 32    | 19693 | 19776  | 1274 | 19434           | 483 |
| 33    | 19926 | 17978  | 1282 | 19436           | 506 |
| 34    | 18150 | 19167  | 1762 | 17875           | 534 |
| 35    | 17893 | 18917  | 1507 | 17639           | 530 |
| 36    | 21412 | 18965  | 1444 | 20964           | 513 |
| 37    | 20659 | 19295  | 1675 | 20311           | 501 |
| 38    | 22521 | 22963  | 873  | 22466           | 435 |
| 39    | 20776 | 21216  | 1625 | 20571           | 494 |
| 40    | 19961 | 21533  | 1543 | 19827           | 510 |
| 41    | 24641 | 22013  | 1559 | 24327           | 530 |
| 42    | 23757 | 26746  | 1677 | 23853           | 531 |
| 43    | 19257 | 19968  | 1597 | 19017           | 509 |
| 44    | 21924 | 21733  | 1828 | 21712           | 496 |
| 45    | 21572 | 20727  | 1440 | 21285           | 484 |
| 46    | 24438 | 25447  | 2051 | 24388           | 516 |
| 47    | 24394 | 25463  | 1415 | 24401           | 498 |
| 48    | 22688 | 24251  | 1500 | 22665           | 499 |
| 49    | 24752 | 24265  | 767  | 24559           | 427 |
| 50    | 31018 | 30788  | 1891 | 31159           | 655 |
| 51    | 24966 | 23744  | 2263 | 24830           | 526 |



Table 2.4. Average Relative Bias, Average Squared Relative Bias, Average Absolute Bias, and Average Squared Deviations of the Estimates.

|                     | Sample Median | HB <sup>1</sup> |
|---------------------|---------------|-----------------|
| Aver. Rel. Bias     | 0.04984       | 0.00980         |
| Aver. Sq. Rel. Bias | 0.00340       | 0.00014         |
| Aver. Abs. Bias     | 1090.41       | 208.45          |
| Aver. Sq. Dev.      | 1631203.47    | 60782.92        |

## CHAPTER 3

### MULTIVARIATE ESTIMATION OF MEANS IN TIME DOMAIN

#### 3.1 Introduction

Periodic sample surveys usually provide sample-based estimates for several parameters. The inter-relationship among the parameters can play an important role in improving the small area estimators. Fay (1987) demonstrated that, even when one parameter of a vector is of interest, estimating the joint vector may improve the estimation of a given parameter compared to restricting attention only to the single parameter. In particular, he suggested improving on estimation of median incomes for four-person families discussed in the previous chapters by including information regarding the median incomes of three- and five- person families as well because of a strong statistical relationship among the three sets. While the HB model of Datta, Fay and Ghosh (1991) involved bivariate data, their analysis was restricted to a fixed point of time.

The univariate HB model considered in Chapter 2 can be extended to the multivariate problems simply by taking the observations at each time as vectors rather than scalars. In Section 3.2 we present a general framework for multivariate HB analysis when the error variance-covariance matrices are known for all time. This section also provides a methodology for obtaining marginal posterior densities through Gibbs sampling algorithm. We revisit the problem of median income estimation in Section 3.3. We find the median income estimates based on the model given in Section 3.2

for the period 1979–1989 for all the states. We have also compared the estimates for the year 1989 with the corresponding estimates obtained in Section 2.4

### 3.2 The General Multivariate Hierarchical Bayes Model

Suppose, based on a given sample at time  $j$ , ( $j = 1, \dots, t$ ) that  $\mathbf{Y}_{ij} = (Y_{ij1}, \dots, Y_{ijs})^T$  is a  $s$ -dimensional column vector of sample survey estimators of some characteristics  $\boldsymbol{\theta}_{ij} = (\theta_{ij1}, \dots, \theta_{ijs})^T$ , for the  $i^{th}$  small area ( $i = 1, \dots, m$ ). As in Chapter 2, our objective is to optimally estimate a linear function  $\sum_{i=1}^m \sum_{j=1}^t \mathbf{l}_{ij}^T \boldsymbol{\theta}_{ij}$ , using all the data available upto time  $t$ . Consider the following HB model:

- I.  $\mathbf{Y}_{ij} | \boldsymbol{\theta}_{ij} \stackrel{ind}{\sim} N(\boldsymbol{\theta}_{ij}, \mathbf{V}_{ij})$  ( $i = 1, \dots, m; j = 1, \dots, t$ );
- II.  $\boldsymbol{\theta}_{ij} | \boldsymbol{\alpha}, \mathbf{b}_j, \boldsymbol{\psi}_j \stackrel{ind}{\sim} N(\mathbf{X}_{ij}\boldsymbol{\alpha} + \mathbf{Z}_{ij}\mathbf{b}_j, \boldsymbol{\psi}_j)$  ( $i = 1, \dots, m; j = 1, \dots, t$ );
- III.  $\mathbf{b}_j | \mathbf{b}_{j-1}, \mathbf{W} \stackrel{ind}{\sim} N(\mathbf{H}_j\mathbf{b}_{j-1}, \mathbf{W})$  ( $j = 1, \dots, t$ );
- IV. Marginally  $\boldsymbol{\alpha}, \boldsymbol{\psi}_1, \dots, \boldsymbol{\psi}_t$ , and  $\mathbf{W}$  are mutually independent with
 
$$\boldsymbol{\alpha} \sim \text{Uniform}(\mathbb{R}^p),$$

$$\boldsymbol{\psi}_j \sim \text{inverse Wishart}(\mathbf{S}_j, k_j),$$

$$\mathbf{W} \sim \text{inverse Wishart}(\mathbf{S}_0, k_0).$$

In (II),  $\mathbf{X}_{ij}$  ( $s \times p$ ) and  $\mathbf{Z}_{ij}$  ( $s \times p$ ) are known design matrices.

In (IV), we shall allow the possibility of diffuse priors for  $\boldsymbol{\psi}_j$  or  $\mathbf{W}$ , e.g.  $\mathbf{S}_j = \mathbf{0}$ , for ( $j = 1, \dots, t$ ),  $\mathbf{S}_0 = \mathbf{0}$ .

The hierarchical structure of the model implies that

- (i)  $\alpha|y, \theta, b_1, \dots, b_t, \psi_1, \dots, \psi_t, W \sim \alpha|\theta, b_1, \dots, b_t, \psi_1, \dots, \psi_t$
- (iia)  $b_j|y, \theta, \alpha, b_{j' \neq j}, \psi_1, \dots, \psi_t, W \sim b_j|\theta, \alpha, b_{j-1}, b_{j+1}, \psi_j, W$   
( $j, j' = 1, \dots, t-1$ )
- (iib)  $b_t|y, \theta, \alpha, b_1, \dots, b_{t-1}, \psi_1, \dots, \psi_t, W \sim b_t|\theta, \alpha, b_{t-1}, \psi_t, W$
- (iii)  $\psi_j|y, \theta, \alpha, b_1, \dots, b_t, W \sim \psi_j|\alpha, \theta, b_1, \dots, b_t$  ( $j = 1, \dots, t$ )
- (iv)  $W|y, \theta, b_1, \dots, b_t, \psi_1, \dots, \psi_t \sim W|b_1, \dots, b_t$
- (v)  $\theta_{ij}|y, b_1, \dots, b_t, \psi_1, \dots, \psi_t, W \sim \theta_{ij}|Y_{ij}, \psi_j, b_j, \alpha$   
( $i = 1, \dots, m; j = 1, \dots, t$ )

Formal derivation of the complete conditional distributions is now straightforward and is omitted. The complete conditional distributions are given by

$$\begin{aligned}
 (i) \quad & \alpha|y, \theta, b_1, \dots, b_t, \psi_1, \dots, \psi_t, W \\
 & \sim N_p \left[ \left( \sum_{i=1}^m \sum_{j=1}^t X_{ij}^T \psi_j^{-1} X_{ij} \right)^{-1} \sum_{i=1}^m \sum_{j=1}^t X_{ij}^T \psi_j^{-1} (\theta_{ij} - Z_{ij}^T b_j), \right. \\
 & \quad \left. \left( \sum_{i=1}^m \sum_{j=1}^t X_{ij}^T \psi_j^{-1} X_{ij} \right)^{-1} \right]
 \end{aligned}$$

(iia) For ( $j = 1, \dots, t-1$ ):

$$\begin{aligned}
 & b_j|y, \theta, \alpha, b_1, \dots, b_{j-1}, b_{j+1}, \dots, b_t, \psi_1, \dots, \psi_t, W \\
 & \sim N_q \left[ \left( \sum_{i=1}^m Z_{ij}^T \psi_j^{-1} Z_{ij} + H_{j+1}^T W^{-1} H_{j+1} + W^{-1} \right)^{-1} \right. \\
 & \quad \times \left\{ \sum_{i=1}^m Z_{ij}^T \psi_j^{-1} (\theta_{ij} - X_{ij} \alpha) \right. \\
 & \quad \left. + H_{j+1}^T W^{-1} b_{j+1} + W^{-1} H_j b_{j-1} \right\}, \\
 & \quad \left. \left( \sum_{i=1}^m Z_{ij}^T \psi_j^{-1} Z_{ij} + H_{j+1}^T W^{-1} H_{j+1} + W^{-1} \right)^{-1} \right]
 \end{aligned}$$

$$(iib) \quad \mathbf{b}_t | \mathbf{y}, \boldsymbol{\theta}, \boldsymbol{\alpha}, \mathbf{b}_1, \dots, \mathbf{b}_t, \boldsymbol{\psi}_1, \dots, \boldsymbol{\psi}_t, \mathbf{W}$$

$$\begin{aligned} & \sim N_q \left[ \left( \sum_{i=1}^m \mathbf{Z}_{it}^T \boldsymbol{\psi}_t^{-1} \mathbf{Z}_{it} + \mathbf{W}^{-1} \right)^{-1} \right. \\ & \quad \times \left\{ \sum_{i=1}^m \mathbf{Z}_{it}^T \boldsymbol{\psi}_t^{-1} (\boldsymbol{\theta}_{it} - \mathbf{X}_{it} \boldsymbol{\alpha}) + \mathbf{W}^{-1} \mathbf{H}_t \mathbf{b}_{t-1} \right\}, \\ & \quad \left. \left( \sum_{i=1}^m \mathbf{Z}_{it}^T \boldsymbol{\psi}_t^{-1} \mathbf{Z}_{it} + \mathbf{W}^{-1} \right)^{-1} \right] \end{aligned}$$

$$(iii) \quad \boldsymbol{\psi}_1, \dots, \boldsymbol{\psi}_t | \mathbf{y}, \boldsymbol{\theta}, \mathbf{b}_1, \dots, \mathbf{b}_t, \mathbf{W}$$

$$\begin{aligned} & \stackrel{ind}{\sim} \text{Inverse Wishart} [\mathbf{S}_j + \sum_{i=1}^m (\boldsymbol{\theta}_{ij} - \mathbf{X}_{ij} \boldsymbol{\alpha} - \mathbf{Z}_{ij} \mathbf{b}_j) \\ & \quad \times (\boldsymbol{\theta}_{ij} - \mathbf{X}_{ij} \boldsymbol{\alpha} - \mathbf{Z}_{ij} \mathbf{b}_j)^T, k_j + m] \end{aligned}$$

$$(iv) \quad \mathbf{W} | \mathbf{y}, \boldsymbol{\theta}, \mathbf{b}_1, \dots, \mathbf{b}_t, \boldsymbol{\psi}_1, \dots, \boldsymbol{\psi}_t$$

$$\sim \text{Inverse Wishart} \left( \mathbf{S}_0 + \sum_{j=1}^t (\mathbf{b}_j - \mathbf{H}_j \mathbf{b}_{j-1}) (\mathbf{b}_j - \mathbf{H}_j \mathbf{b}_{j-1})^T, k_0 + t \right)$$

$$(v) \quad \boldsymbol{\theta}_{ij} | \mathbf{y}, \mathbf{b}_1, \dots, \mathbf{b}_t, \boldsymbol{\psi}_1, \dots, \boldsymbol{\psi}_t, \mathbf{W}$$

$$\begin{aligned} & \sim N \left[ \left( \mathbf{V}_{ij}^{-1} + \boldsymbol{\psi}_j^{-1} \right)^{-1} \left( \mathbf{V}_{ij}^{-1} \mathbf{y}_{ij} + \boldsymbol{\psi}_j^{-1} (\mathbf{X}_{ij} \boldsymbol{\alpha} + \mathbf{Z}_{ij} \mathbf{b}_j) \right), \right. \\ & \quad \left. \left( \mathbf{V}_{ij}^{-1} + \boldsymbol{\psi}_j^{-1} \right)^{-1} \right] \end{aligned}$$

A Rao–Blackwellized sample based density estimate of  $\theta_{ij}$  given  $\mathbf{y}$  based on simulated samples from the above conditional distributions can be approximated by

$$\frac{1}{G} \sum_{g=1}^G f \left( \theta_{ij} | \mathbf{y}, \boldsymbol{\alpha} = \boldsymbol{\alpha}_g^{(t)}, \mathbf{W} = \mathbf{W}_g^{(t)}, \mathbf{b}_j = \mathbf{b}_{jg}, \boldsymbol{\psi}_j = \boldsymbol{\psi}_{jg}^{(t)}, j = 1, \dots, t \right)$$

### 3.3 An Illustrative Application to Survey data

In this section, we will illustrate the implementation of multivariate HB model described in Section 3.2 . We again consider the problem of estimation of median



income for four-person families by state. Fay (1987) suggested improving on estimation of median incomes for four-person families by including information regarding the median incomes of three- and five- person families as well because of a strong statistical relationship among the three sets. He advocated the use of a multivariate empirical Bayes (EB) or variance-components approach for the composite estimation of median incomes of four-person families by state. Datta, Fay and Ghosh (1991) adopted a bivariate hierarchical linear model for estimating the statewide median incomes of four-person families. In their procedure, one component was the median income of four-person families, while the other component was a convex combination of the median incomes of five- and three-person families. The analysis required evaluation of 3-dimensional integrals. Instead, we will use Monte-carlo integration via Gibbs sampling algorithm. In our analysis, we will be considering as the basic data 51 three component vectors corresponding to the median incomes of four-, three-, and five person families for the years 1979–1989 (except for 1980). Our target is to estimate statewide median incomes of four-person families, although our procedure furnishes estimates of median incomes of three- and five person families as well. Note that direct numerical integration in a trivariate set up will require evaluation of six-dimensional integrals, which is fairly complicated and often unreliable. The Gibbs sampling technique avoids such problems. Our analysis rests heavily on the implementation of the Gibbs sampling algorithm.

We consider as the basic data the three component vector  $\mathbf{Y}_{ij} = (Y_{ij1}, Y_{ij2}, Y_{ij3})^T$ ,  $i = 1, \dots, 51; j = 1, \dots, 10$ , where  $Y_{ij1}$ ,  $Y_{ij2}$ , and  $Y_{ij3}$  are the sample median incomes of four-, three-, and five-person families in state  $i$  in year  $j$ . The true medians corresponding to  $Y_{ij}$ 's are denoted by  $\theta_{ij}$ 's respectively. Let  $\theta_{ij} = (\theta_{ij1}, \theta_{ij2}, \theta_{ij3})^T$ . We shall also write as  $\mathbf{Y} = (\mathbf{Y}_{11}, \dots, \mathbf{Y}_{51,10})^T$ ,  $\boldsymbol{\theta} = (\boldsymbol{\theta}_{11}, \dots, \boldsymbol{\theta}_{51,10})^T$ .

We would like to introduce few more notations before describing the HB model. Let  $x_{ij11}$  and  $x_{ij12}$  denote respectively the adjusted census median income for the

year  $j$  and the base-year census median income for four-person families in the  $i^{th}$  state ( $i = 1, \dots, 51$ ). Also, let  $x_{ij21}$  and  $x_{ij31}$  denote respectively the adjusted census median incomes for the year  $j$  for three-, and five-person families. The corresponding base-year census median income for three-, and five-person families are denoted by  $x_{ij22}$  and  $x_{ij32}$ . Finally, let

$$\mathbf{X}_{ij} = \begin{pmatrix} 1 & x_{ij11} & x_{ij12} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & x_{ij21} & x_{ij22} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & x_{ij31} & x_{ij32} \end{pmatrix}, \quad (3.3.1)$$

$i = 1, \dots, 51$ ,  $j = 1, \dots, 10$ , and denote by  $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_9)^T$  the vector of regression coefficients. Also,  $\mathbf{b}_j = (b_{j1} \ b_{j2} \ b_{j3})^T$ , ( $j = 1, \dots, 10$ ) is the vector of random componets for the year  $j$ .

The following HB model is used:

- I.  $\mathbf{Y}_{ij} | \boldsymbol{\theta}_{ij} \stackrel{ind}{\sim} N(\boldsymbol{\theta}_{ij}, \mathbf{V}_{ij})$  ( $i = 1, \dots, 51; j = 1, \dots, 10$ );
- II.  $\boldsymbol{\theta}_{ij} | \boldsymbol{\alpha}, \mathbf{b}_j, \boldsymbol{\psi}_j \stackrel{ind}{\sim} N(\mathbf{X}_{ij}\boldsymbol{\alpha} + \mathbf{b}_j, \boldsymbol{\psi}_j)$  ( $i = 1, \dots, 51; j = 1, \dots, 10$ );
- III.  $\mathbf{b}_j | \mathbf{b}_{j-1}, \mathbf{W} \stackrel{ind}{\sim} N(\mathbf{b}_{j-1}, \mathbf{W})$  ( $j = 1, \dots, 10$ );
- IV. Marginally  $\boldsymbol{\alpha}, \boldsymbol{\psi}_1, \dots, \boldsymbol{\psi}_t$ , and  $\mathbf{W}$  are mutually independent with
 
$$\boldsymbol{\alpha} \sim \text{Uniform}(R^9),$$

$$\boldsymbol{\psi}_j \sim \text{inverse Wishart}(\mathbf{S}_j, k_j),$$

$$\mathbf{W} \sim \text{inverse Wishart}(\mathbf{S}_0, k_0).$$

In (II),  $\mathbf{X}_{ij}$  ( $3 \times p$ ) is a known design matrix.

In (IV) we assume diffuse priors for  $\boldsymbol{\psi}_j$  or  $\mathbf{W}$ , e.g.  $\mathbf{S}_j = \mathbf{0}$ , for ( $j = 1, \dots, 10$ ),  $\mathbf{S}_0 = \mathbf{0}$ .

As mentioned earlier, closed-form expressions for  $E(\boldsymbol{\theta}_{ij} | \mathbf{y})$  and  $V(\boldsymbol{\theta}_{ij} | \mathbf{y})$  in such a set-up can not be obtained. Also, it is too complicated to estimate  $E(\boldsymbol{\theta}_{ij} | \mathbf{y})$  and

$V(\theta_{ij}|\mathbf{y})$  through numerical integration. We will be using the Gibbs sampling procedure to simulate samples from the following complete conditional distributions:

$$(i) \quad \alpha|\mathbf{y}, \theta, \mathbf{b}_1, \dots, \mathbf{b}_{10}, \psi_1, \dots, \psi_{10}, \mathbf{W}$$

$$\sim N_p \left[ \left( \sum_{i=1}^{51} \sum_{j=1}^{10} \mathbf{X}_{ij}^T \psi_j^{-1} \mathbf{X}_{ij} \right)^{-1} \sum_{i=1}^{51} \sum_{j=1}^{10} \mathbf{X}_{ij}^T \psi_j^{-1} (\theta_{ij} - \mathbf{b}_j), \right. \\ \left. \left( \sum_{i=1}^{51} \sum_{j=1}^{10} \mathbf{X}_{ij}^T \psi_j^{-1} \mathbf{X}_{ij} \right)^{-1} \right]$$

$$(iia) \quad \text{For } (j = 1, \dots, 9) :$$

$$\mathbf{b}_j|\mathbf{y}, \theta, \alpha, \mathbf{b}_1, \dots, \mathbf{b}_{j-1}, \mathbf{b}_{j+1}, \dots, \mathbf{b}_{10}, \psi_1, \dots, \psi_{10}, \mathbf{W}$$

$$\sim N_3 \left[ \left( 51 \psi_j^{-1} + 2 \mathbf{W}^{-1} \right)^{-1} \right. \\ \times \left\{ \psi_j^{-1} \sum_{i=1}^{51} (\theta_{ij} - \mathbf{X}_{ij} \alpha) \right. \\ \left. + \mathbf{W}^{-1} (\mathbf{b}_{j+1} + \mathbf{b}_{j-1}) \right\}, \\ \left. \left( 51 \psi_j^{-1} + 2 \mathbf{W}^{-1} \right)^{-1} \right]$$

$$(iib) \quad \mathbf{b}_{10}|\mathbf{y}, \theta, \alpha, \mathbf{b}_1, \dots, \mathbf{b}_{10}, \psi_1, \dots, \psi_{10}, \mathbf{W}$$

$$\sim N_3 \left[ \left( 51 \psi_{10}^{-1} + \mathbf{W}^{-1} \right)^{-1} \right. \\ \times \left\{ \psi_{10}^{-1} \sum_{i=1}^{51} (\theta_{i10} - \mathbf{X}_{i10} \alpha) + \mathbf{W}^{-1} \mathbf{b}_9 \right\}, \\ \left. \left( 51 \psi_{10}^{-1} + \mathbf{W}^{-1} \right)^{-1} \right]$$

$$(iii) \quad \psi_1, \dots, \psi_{10}|\mathbf{y}, \theta, \mathbf{b}_1, \dots, \mathbf{b}_{10}, \mathbf{W}$$

$$\stackrel{ind}{\sim} \text{Inverse Wishart} \left[ \mathbf{S}_j + \sum_{i=1}^{10} (\theta_{ij} - \mathbf{X}_{ij} \alpha - \mathbf{b}_j) \right. \\ \left. \times (\theta_{ij} - \mathbf{X}_{ij} \alpha - \mathbf{b}_j)^T, 51 \right]$$

$$(iv) \quad \mathbf{W}|\mathbf{y}, \theta, \mathbf{b}_1, \dots, \mathbf{b}_{10}, \psi_1, \dots, \psi_{10}$$

$$\sim \text{Inverse Wishart} \left( \sum_{j=1}^{10} (\mathbf{b}_j - \mathbf{b}_{j-1})(\mathbf{b}_j - \mathbf{b}_{j-1})^T, 10 \right)$$

$$(v) \quad \theta_{ij} | \mathbf{y}, \mathbf{b}_1, \dots, \mathbf{b}_{10}, \psi_1, \dots, \psi_{10}, \mathbf{W}$$

$$\sim N \left[ \left( \mathbf{V}_{ij}^{-1} + \psi_j^{-1} \right)^{-1} \left( \mathbf{V}_{ij}^{-1} \mathbf{y}_{ij} + \psi_j^{-1} (\mathbf{X}_{ij} \boldsymbol{\alpha} + \mathbf{b}_j) \right), \right. \\ \left. \left( \mathbf{V}_{ij}^{-1} + \psi_j^{-1} \right)^{-1} \right]$$

With all the complete conditional distributions available in closed form, we can now estimate the marginal posterior density of  $\theta_{ij} | \mathbf{y}$  as described earlier in Section 4.2. We find estimates of the four-person families for each year from 1979–1989 (except for 1980) using all the sample estimates available for three-, four-, and five person families since 1979 census (except for 1980) by running 50 iterations of 100–2000 replications each. The necessary convergence is achieved after 500 iterations. We denote these HB estimates by HB<sup>3</sup>. As in Section 2.4, we also consider the non time-series analysis which utilizes only the census median income figures for 1979 and the CPS median income estimates for 1989, and the PCI's for the years 1979 and 1989. In this case, we utilize the data available for three-, four-, and five person families in contrast to using only the information for four-person families as in Section 2.4. We denote these estimates by HB<sup>4</sup>. These estimates are based on a trivariate version of the HB model considered in Datta, Fay, and Ghosh (1991). Table 3.1 provides the sample median incomes as well as the HB<sup>3</sup> and HB<sup>4</sup> estimates. The HB<sup>3</sup> estimates are based on the time series HB described earlier in this section. The standard errors associated with these estimates are given in Table 3.2. Again in conformity with the results in the case of univariate analysis in Chapter 2, the time domain HB procedure results in substantial reduction of the standard errors when compared to the sample estimates. Also, the standard errors associated with the HB<sup>3</sup> estimates are lower than the standard errors of the HB<sup>4</sup> estimates for each state.

The median income estimates along with their standard errors for the year 1989 are provided in Table 3.3. Also, as in Section 2.4, we compare the retrospective estimates obtained for 1979 by using the time series model against the 1980 census,



treating the latter as the “truth”. We also compute the criteria average relative bias, average squared relative bias, average absolute bias, and average squared deviation (as defined in Section 2.4) for the HB<sup>3</sup> estimates. These quantities are provided in Table 3.4. Again, the HB<sup>3</sup> estimates are clearly superior to the sample median estimates for each state. In terms of the average relative bias, the HB<sup>3</sup> estimates improve on the sample medians by 87.62 %. Using the criterion of average squared relative bias, the corresponding percentage improvement is 98.24. Similarly, for average absolute bias and average squared deviations as the criteria, the respective percentage improvement of the HB<sup>3</sup> estimates over the sample medians are 87.69 and 98.28. Thus time series HB procedure improve tremendously on the sample estimates.

We also compare the HB<sup>3</sup> estimates for the year 1979 based on the trivariate time series HB model with the HB<sup>1</sup> estimates based on univariate analysis in Chapter 2. In terms of the average relative bias, the HB<sup>3</sup> estimates improve on the HB<sup>1</sup> estimates by 37.04 %. Using the criterion of average squared relative bias, the corresponding percentage improvement is 57.14. Similarly, for average absolute bias and average squared deviations as the criteria, the respective percentage improvement of the HB<sup>3</sup> estimates over the HB<sup>1</sup> estimates are 33.67 and 53.94. Thus, the trivariate time series HB model captures the inter-relationships between the median income of three-, four- and five-person families to improve the median income estimates of four-person families.



Table 3.1. Median Income for 4-Person Family by State in 1989

(In Dollars )

| State | Sample<br>Median | HB <sup>3</sup> | HB <sup>4</sup> |
|-------|------------------|-----------------|-----------------|
| 1     | 37700            | 37826           | 38172           |
| 2     | 52720            | 49130           | 48988           |
| 3     | 41489            | 40889           | 41118           |
| 4     | 55729            | 53372           | 52986           |
| 5     | 47106            | 43723           | 43578           |
| 6     | 55048            | 54755           | 54140           |
| 7     | 43434            | 43760           | 43513           |
| 8     | 53128            | 53756           | 53076           |
| 9     | 40990            | 40582           | 40508           |
| 10    | 44241            | 42408           | 42326           |
| 11    | 40297            | 38250           | 38121           |
| 12    | 42590            | 42732           | 42360           |
| 13    | 43260            | 43405           | 43090           |
| 14    | 42226            | 41033           | 40932           |
| 15    | 45432            | 43043           | 42748           |
| 16    | 38711            | 37100           | 37122           |
| 17    | 39232            | 38512           | 38491           |
| 18    | 36966            | 35498           | 35790           |
| 19    | 32610            | 33137           | 33514           |
| 20    | 40894            | 38846           | 38988           |
| 21    | 42781            | 38934           | 38990           |
| 22    | 39819            | 43068           | 42751           |
| 23    | 48252            | 51029           | 50229           |
| 24    | 36587            | 39918           | 39842           |
| 25    | 49718            | 46250           | 46019           |

Table 3.1. continued

| State | Sample<br>Median | HB <sup>3</sup> | HB <sup>4</sup> |
|-------|------------------|-----------------|-----------------|
| 26    | 31604            | 31102           | 31346           |
| 27    | 37423            | 38078           | 38174           |
| 28    | 31834            | 35686           | 35865           |
| 29    | 45876            | 40665           | 40830           |
| 30    | 37194            | 36952           | 36919           |
| 31    | 32186            | 33722           | 33976           |
| 32    | 31327            | 33911           | 34048           |
| 33    | 32245            | 34192           | 34365           |
| 34    | 31330            | 31902           | 32284           |
| 35    | 33317            | 31516           | 32021           |
| 36    | 36219            | 34409           | 34543           |
| 37    | 40358            | 34252           | 34551           |
| 38    | 32679            | 34226           | 34137           |
| 39    | 29586            | 32865           | 32944           |
| 40    | 32227            | 33112           | 33319           |
| 41    | 38931            | 37177           | 36842           |
| 42    | 36732            | 40034           | 39727           |
| 43    | 29761            | 30098           | 30441           |
| 44    | 38969            | 38132           | 38100           |
| 45    | 37938            | 36890           | 36953           |
| 46    | 34579            | 38654           | 38221           |
| 47    | 42392            | 42561           | 42316           |
| 48    | 41419            | 39475           | 39404           |
| 49    | 41121            | 42787           | 42421           |
| 50    | 41379            | 49151           | 47671           |
| 51    | 44999            | 45806           | 45262           |

Table 3.2. Estimated Standard Errors of Median Income Estimates  
for 4-Person Family by State in 1989

(In Dollars )

| State | Sample<br>Median | HB <sup>3</sup> | HB <sup>4</sup> |
|-------|------------------|-----------------|-----------------|
| 1     | 3230             | 1867            | 2051            |
| 2     | 4132             | 2129            | 2281            |
| 3     | 2848             | 1825            | 1925            |
| 4     | 2452             | 1710            | 1848            |
| 5     | 4512             | 2065            | 2168            |
| 6     | 3893             | 2291            | 2399            |
| 7     | 1512             | 1167            | 1231            |
| 8     | 1933             | 1389            | 1545            |
| 9     | 1584             | 1157            | 1194            |
| 10    | 1845             | 1350            | 1399            |
| 11    | 2970             | 1774            | 1830            |
| 12    | 1763             | 1272            | 1342            |
| 13    | 1597             | 1225            | 1277            |
| 14    | 2310             | 1581            | 1630            |
| 15    | 3251             | 1865            | 1925            |
| 16    | 2258             | 1501            | 1560            |
| 17    | 3677             | 1875            | 1928            |
| 18    | 2249             | 1537            | 1615            |
| 19    | 2130             | 1494            | 1586            |
| 20    | 2502             | 1687            | 1737            |
| 21    | 3359             | 1952            | 2016            |
| 22    | 3699             | 1972            | 2024            |
| 23    | 4665             | 2214            | 2355            |
| 24    | 4645             | 2235            | 2295            |
| 25    | 4349             | 2017            | 2104            |

Table 3.2. continued

| State | Sample<br>Median | HB <sup>3</sup> | HB <sup>4</sup> |
|-------|------------------|-----------------|-----------------|
| 26    | 2278             | 1515            | 1594            |
| 27    | 1513             | 1140            | 1194            |
| 28    | 2965             | 1766            | 1842            |
| 29    | 4385             | 2049            | 2148            |
| 30    | 1397             | 1113            | 1154            |
| 31    | 2885             | 1774            | 1839            |
| 32    | 2446             | 1702            | 1761            |
| 33    | 2690             | 1706            | 1768            |
| 34    | 2392             | 1614            | 1701            |
| 35    | 2435             | 1587            | 1704            |
| 36    | 3314             | 1841            | 1931            |
| 37    | 3923             | 2031            | 2135            |
| 38    | 1552             | 1287            | 1292            |
| 39    | 2557             | 1693            | 1731            |
| 40    | 2268             | 1516            | 1573            |
| 41    | 3516             | 2030            | 2280            |
| 42    | 3261             | 1897            | 1944            |
| 43    | 2811             | 1661            | 1773            |
| 44    | 5003             | 2022            | 2089            |
| 45    | 3209             | 1745            | 1809            |
| 46    | 2863             | 1863            | 1914            |
| 47    | 2525             | 1679            | 1749            |
| 48    | 3056             | 1821            | 1881            |
| 49    | 1586             | 1216            | 1235            |
| 50    | 3637             | 2260            | 2607            |
| 51    | 4805             | 2144            | 2252            |

Table 3.3. Median Income Estimates for 4-Person Family by State in 1979

(In Dollars )

| State | Truth | Sample |      | HB <sup>3</sup> |     |
|-------|-------|--------|------|-----------------|-----|
|       |       | Median | SE   | Median          | SE  |
| 1     | 18319 | 18084  | 1533 | 18206           | 511 |
| 2     | 22027 | 23129  | 1647 | 22032           | 481 |
| 3     | 19424 | 18438  | 1642 | 19246           | 498 |
| 4     | 23772 | 25324  | 1414 | 23808           | 467 |
| 5     | 22107 | 23597  | 2409 | 22094           | 499 |
| 6     | 25712 | 25037  | 1431 | 25631           | 493 |
| 7     | 22669 | 21852  | 766  | 22414           | 395 |
| 8     | 26014 | 25486  | 1264 | 25938           | 482 |
| 9     | 22266 | 23326  | 878  | 22440           | 453 |
| 10    | 23279 | 22397  | 843  | 23061           | 412 |
| 11    | 23014 | 22716  | 1041 | 22819           | 449 |
| 12    | 25410 | 24294  | 1025 | 25167           | 459 |
| 13    | 25111 | 24103  | 1006 | 24963           | 454 |
| 14    | 23320 | 23270  | 1202 | 23322           | 471 |
| 15    | 24044 | 25137  | 1571 | 23928           | 526 |
| 16    | 22351 | 23851  | 1592 | 22375           | 477 |
| 17    | 21891 | 20743  | 1255 | 21695           | 456 |
| 18    | 20511 | 18567  | 1535 | 20211           | 491 |
| 19    | 18674 | 19826  | 1588 | 18621           | 506 |
| 20    | 21438 | 22603  | 1802 | 21395           | 482 |
| 21    | 22127 | 22014  | 1586 | 22025           | 466 |
| 22    | 23627 | 21860  | 1900 | 23476           | 493 |
| 23    | 26203 | 26235  | 1722 | 26202           | 517 |
| 24    | 21862 | 20232  | 3801 | 21737           | 507 |
| 25    | 22757 | 24160  | 1418 | 22866           | 486 |



Table 3.3. continued

| State | Truth | Sample |      | HB <sup>3</sup> |     |
|-------|-------|--------|------|-----------------|-----|
|       |       | Median | SE   | Median          | SE  |
| 26    | 20214 | 18274  | 1380 | 19890           | 496 |
| 27    | 19772 | 20296  | 1012 | 19702           | 445 |
| 28    | 19944 | 19282  | 1795 | 19757           | 502 |
| 29    | 20668 | 22687  | 1196 | 20824           | 489 |
| 30    | 21086 | 19675  | 1042 | 20717           | 464 |
| 31    | 19685 | 18657  | 1285 | 19539           | 484 |
| 32    | 19693 | 19776  | 1274 | 19532           | 471 |
| 33    | 19926 | 17978  | 1282 | 19541           | 510 |
| 34    | 18150 | 19167  | 1762 | 18106           | 517 |
| 35    | 17893 | 18917  | 1507 | 17844           | 517 |
| 36    | 21412 | 18965  | 1444 | 21084           | 491 |
| 37    | 20659 | 19295  | 1675 | 20426           | 489 |
| 38    | 22521 | 22963  | 873  | 22501           | 416 |
| 39    | 20776 | 21216  | 1625 | 20646           | 482 |
| 40    | 19961 | 21533  | 1543 | 20008           | 500 |
| 41    | 24641 | 22013  | 1559 | 24412           | 505 |
| 42    | 23757 | 26746  | 1677 | 23944           | 518 |
| 43    | 19257 | 19968  | 1597 | 19167           | 496 |
| 44    | 21924 | 21733  | 1828 | 21861           | 489 |
| 45    | 21572 | 20727  | 1440 | 21366           | 469 |
| 46    | 24438 | 25447  | 2051 | 24416           | 498 |
| 47    | 24394 | 25463  | 1415 | 24399           | 477 |
| 48    | 22688 | 24251  | 1500 | 22720           | 481 |
| 49    | 24752 | 24265  | 767  | 24518           | 418 |
| 50    | 31018 | 30788  | 1891 | 30937           | 660 |
| 51    | 24966 | 23744  | 2263 | 24873           | 508 |

Table 3.4. Average Relative Bias, Average Squared Relative Bias, Average Absolute Bias, and Average Squared Deviations of the Estimates.

|                     | Sample Median | HB <sup>3</sup> |
|---------------------|---------------|-----------------|
| Aver. Rel. Bias     | 0.04984       | 0.00617         |
| Aver. Sq. Rel. Bias | 0.00340       | 0.00006         |
| Aver. Abs. Bias     | 1090.41       | 134.27          |
| Aver. Sq. Dev.      | 1631203.47    | 27996.90        |

## CHAPTER 4

### HIERARCHICAL BAYES PREDICTION IN FINITE POPULATIONS

#### 4.1 Introduction

In Chapter 2 and Chapter 3, we introduced HB procedures for small domain estimation based on data from repeated surveys from an infinite population. In such models, we assumed that the observations are available only at the aggregate level. There we derived the HB estimators for the current as well as previous time points, simultaneously for all the small area population means.

In this Chapter, we consider the problem of small domain estimation in the context of finite population sampling. In the finite population setup, we assume that there are  $m$  strata, the  $i^{th}$  stratum at time  $j$  contains a finite number of units  $N_{ij}$ , and the unit level information on the characteristic of interest at time  $j$ , ( $j = 1, \dots, t$ ) is available through a sequence of random sample surveys of the given strata populations. We are interested in predicting certain parameters (such as finite population mean or total or change in mean over two distinct time points for each stratum) of the strata populations which are changing over time.

We describe a general HB approach in Section 4.2. There we consider the most general situation where all the variance components are unknown. Some prior distributions are assigned to the vector of regression coefficients and the variance components. This section provides the predictive distribution of unobserved population

units at all the time points given the time series data on the sampled units. The computation of the estimation of population means simultaneously for several small areas at all the time points through the implementation of the Gibbs sampling algorithm (as described in Section 2.2.1) have also been discussed in this section.

In Section 4.3, we consider a special case of the general HB model in Section 4.2. We assume that  $\lambda$ , the vector of ratios of variance components is known. In such a case, it is possible to find the posterior mean vector and posterior variance-covariance matrix corresponding to the characteristic vector of non-sampled units at all the time points, in a closed form. In subsequent sections 4.4–4.5, we assume this special case of the HB model. The corresponding HB predictor as obtained in Section 4.3 is shown to possess some interesting optimal properties. Datta and Ghosh (1991) also assumed known ratios of the variance components in a special case of their general HB approach for small area estimation. The HB predictors in their special case have been shown to be the best within the class of all linear unbiased predictors. For a class of spherically symmetric distributions, their HB predictors are also optimal within the class of all unbiased predictors. In Datta and Ghosh (1990) these HB predictors dominate stochastically the linear unbiased predictors in the sense of Hwang (1985). Furthermore, under a suitable group of transformations, Datta and Ghosh (1990–91) showed these predictors to be the “best” within the class of all equivariant predictors for elliptically symmetric distributions. In the set up of Datta and Ghosh, the HB predictor borrowed strength from other small areas. In our set-up, as we have mentioned earlier in Chapter 2, the HB predictor borrows strength over time as well in addition to borrowing strength from other small areas. Our HB predictors as derived in Section 4.3 based on a time series HB model, possesses all the optimal properties listed above. In establishing this in Sections 4.4–4.6, we have followed the approach as in Datta and Ghosh (1990–91). In Section 4.4, it is shown that the HB predictors of section 4.3 are best unbiased predictors within the class of



all unbiased predictors. Besides, it is shown that these predictors are BLUPs. Section 4.5 establishes the stochastic dominance (cf. Hwang, 1985) of these predictors over the LUPs for elliptically symmetric distributions. Finally, in Section 4.6, we have shown that these predictors are also best equivariant predictors for both the matrix loss (or standardized quadratic loss) under suitable group of transformations for a class of elliptically symmetric distributions.

## 4.2 Description of the Hierarchical Bayes Model

In this section, we consider a hierarchical model for prediction of the characteristic of interest (like the population total or mean for each small area or domain) in the context of finite population sampling. Let  $Y_{ijk}$  denote some characteristic of interest associated with the  $k^{th}$  unit at time  $j$  in the  $i^{th}$  small area ( $k = 1, \dots, N_{ij}; j = 1, \dots, t; i = 1, \dots, m$ ). For notational simplicity, we assume that the characteristic of interest is a scalar. The results can be easily extended to the case when  $Y_{ijk}$  is a vector. Let

$$Y_{ij}^{(1)} = (Y_{ij1}, \dots, Y_{ij, n_{ij}})^T$$

$$Y_{ij}^{(2)} = (Y_{ij, n_{ij}+1}, \dots, Y_{ij, N_{ij}})^T$$

The superscripts (1) and (2) are used to denote the sampled and non-sampled units in the  $i^{th}$  small area at time  $j$ , respectively.  $N_{ij}$  is the population size of the  $i^{th}$  small area at time  $j$  from which a sample of size  $n_{ij}$  is taken ( $1 \leq i \leq m; 1 \leq j \leq t$ ).  $X_{ij}^{(1)}$  ( $n_{ij} \times p$ ) and  $Z_{ij}^{(1)}$  ( $n_{ij} \times q$ ) known design matrices correspond to the sampled units.  $X_{ij}^{(2)}$  ( $(N_{ij} - n_{ij}) \times p$ ) and  $Z_{ij}^{(2)}$  ( $(N_{ij} - n_{ij}) \times q$ ) correspond to the unsampled units.

We consider the following Bayesian model:



$$\text{I. } \left( \begin{array}{c} Y_{ij}^{(1)} \\ Y_{ij}^{(2)} \end{array} \right) \middle| \left( \begin{array}{c} \theta_{ij}^{(1)} \\ \theta_{ij}^{(2)} \end{array} \right), \sigma^2, \lambda \stackrel{\text{ind}}{\sim} N \left( \left( \begin{array}{c} \theta_{ij}^{(1)} \\ \theta_{ij}^{(2)} \end{array} \right), \sigma^2 V_{ij}(\lambda) \right);$$

$$(i = 1, \dots, m; j = 1, \dots, t)$$

$$\text{II. } \left( \begin{array}{c} \theta_{ij}^{(1)} \\ \theta_{ij}^{(2)} \end{array} \right) \middle| \alpha, b_j, \sigma^2, \lambda \stackrel{\text{ind}}{\sim} N \left( \left( \begin{array}{c} X_{ij}^{(1)} \\ X_{ij}^{(2)} \end{array} \right) \alpha + \left( \begin{array}{c} Z_{ij}^{(1)} \\ Z_{ij}^{(2)} \end{array} \right) b_j, \sigma^2 \psi_j(\lambda) \right);$$

$$(i = 1, \dots, m; j = 1, \dots, t)$$

$$\text{III. } b_j | b_{j-1}, W, \lambda \stackrel{\text{ind}}{\sim} N(H_j b_{j-1}, \sigma^2 W(\lambda)); (j = 1, \dots, t)$$

We assume that  $V_{ij}(\lambda)$ ,  $\psi_j(\lambda)$ ,  $(i = 1, \dots, m; j = 1, \dots, t)$ , and  $W(\lambda)$  are structurally known except possibly for some unknown  $\lambda = (\lambda_1, \dots, \lambda_s)$ .  $H_j (q \times q)$ ,  $(j = 1, \dots, t)$  are known matrices.  $b_0$  is assumed to be known. Without any loss of generality, we assume  $b_0 = 0$ .

Before writing the stages I–III of the above model in a compact form, we need to introduce a few notations. We denote a  $u \times u$  identity matrix by  $I_u$ .  $\Phi$  denotes a null matrix of suitable order.

We define

$$C_t^{(1)} = \left( C_{1t}^{(1)T}, C_{2t}^{(1)T}, \dots, C_{mt}^{(1)T} \right)^T$$

$$C_{it}^{(1)} = \left( Y_{i1}^{(1)T}, Y_{i2}^{(1)T}, \dots, Y_{it}^{(1)T} \right)^T; (i = 1, \dots, m)$$

$$C_t^{(2)} = \left( C_{1t}^{(2)T}, C_{2t}^{(2)T}, \dots, C_{mt}^{(2)T} \right)^T$$

$$C_{it}^{(2)} = \left( Y_{i1}^{(2)T}, Y_{i2}^{(2)T}, \dots, Y_{it}^{(2)T} \right)^T; (i = 1, \dots, m)$$

$$X_i^{(1)} = \left( X_{i1}^{(1)T}, X_{i2}^{(1)T}, \dots, X_{it}^{(1)T} \right)^T; (i = 1, \dots, m)$$

$$X_i^{(2)} = \left( X_{i1}^{(2)T}, X_{i2}^{(2)T}, \dots, X_{it}^{(2)T} \right)^T; (i = 1, \dots, m)$$

$$A_t^{(1)} = \left( X_1^{(1)T}, X_2^{(1)T}, \dots, X_m^{(1)T} \right)^T$$

$$A_t^{(2)} = \left( X_1^{(2)T}, X_2^{(2)T}, \dots, X_m^{(2)T} \right)^T$$

$$\begin{aligned} U_{jl} &= \prod_{k=l}^j H_k, \quad l = 1, 2, \dots, j; \quad j = 1, \dots, t \\ &= I_q \text{ for } l = j + 1, \quad j = 1, \dots, t \end{aligned}$$

$$B_{it}^{(1)} = \begin{bmatrix} Z_{i1}^{(1)} & \Phi & \dots & \Phi \\ Z_{i2}^{(1)} U_{22} & Z_{i2}^{(1)} & \dots & \Phi \\ \vdots & \vdots & & \vdots \\ Z_{it}^{(1)} U_{t2} & Z_{it}^{(1)} U_{t3} & \dots & Z_{it}^{(1)} \end{bmatrix}$$

$$B_{it}^{(2)} = \begin{bmatrix} Z_{i1}^{(2)} & \Phi & \dots & \Phi \\ Z_{i2}^{(2)} U_{22} & Z_{i2}^{(2)} & \dots & \Phi \\ \vdots & \vdots & & \vdots \\ Z_{it}^{(2)} U_{t2} & Z_{it}^{(2)} U_{t3} & \dots & Z_{it}^{(2)} \end{bmatrix}$$

$$B_t^{(1)} = \begin{bmatrix} B_{1t}^{(1)} \\ B_{2t}^{(1)} \\ \vdots \\ B_{mt}^{(1)} \end{bmatrix}, \quad B_t^{(2)} = \begin{bmatrix} B_{1t}^{(2)} \\ B_{2t}^{(2)} \\ \vdots \\ B_{mt}^{(2)} \end{bmatrix}$$

$$V_t = \text{Block Diag} (V_{11}, \dots, V_{1t}, V_{21}, \dots, V_{2t}, \dots, V_{m1}, \dots, V_{mt})$$

$$\psi_t = \text{Block Diag} (\psi_1, \dots, \psi_t)$$

We now write the model as a compact linear mixed model which incorporates all the information up to and including time  $t$  for all the small areas. We have, without any loss of generality, listed the sampled units first.

$$\begin{pmatrix} C_t^{(1)} \\ C_t^{(2)} \end{pmatrix} = \begin{pmatrix} A_t^{(1)} \\ A_t^{(2)} \end{pmatrix} \alpha + \begin{pmatrix} B_t^{(1)} \\ B_t^{(2)} \end{pmatrix} \omega + \begin{pmatrix} \eta^{(1)} \\ \eta^{(2)} \end{pmatrix} \quad (4.2.1)$$

where

$$\begin{aligned} \omega &\sim N(0, \sigma^2 (I_t \otimes W(\lambda))) \\ \eta &= \begin{pmatrix} \eta^{(1)} \\ \eta^{(2)} \end{pmatrix} \\ &\sim N(0, \sigma^2 ((I_m \otimes \psi_t(\lambda)) + V_t(\lambda))) \end{aligned}$$

Writing

$$C_t = \begin{pmatrix} C_t^{(1)} \\ C_t^{(2)} \end{pmatrix}, \quad A_t = \begin{pmatrix} A_t^{(1)} \\ A_t^{(2)} \end{pmatrix}, \quad B_t = \begin{pmatrix} B_t^{(1)} \\ B_t^{(2)} \end{pmatrix}, \quad R = (\sigma^2)^{-1}$$

given  $\alpha, \lambda, \sigma^2$ , and  $r$

$$C_t \sim N(A_t \alpha, r^{-1} \Sigma_t) \quad (4.2.2)$$

where

$$\Sigma_t \equiv \Sigma_t(\lambda) = B_t (I_t \otimes W) B_t^T + (I_m \otimes \Psi_t) + V_t$$

Furthermore, we partition  $\Sigma_t$  as

$$\Sigma_t = \begin{bmatrix} \Sigma_{t11} & \Sigma_{t12} \\ \Sigma_{t21} & \Sigma_{t22} \end{bmatrix}$$

where  $\Sigma_{t11}$  is  $n_t \times n_t$ ,  $\Sigma_{t12}$  ( $= \Sigma_{t12}^T$ ) is  $n_t \times (N_t - n_t)$ , and  $\Sigma_{t22}$  is  $(N_t - n_t) \times (N_t - n_t)$ .

$$n_t = \sum_{i=1}^m \sum_{j=1}^t n_{ij}$$

$$N_t = \sum_{i=1}^m \sum_{j=1}^t N_{ij}$$

We also define

$$\Sigma_{t22.1} = \Sigma_{t22} - \Sigma_{t21} \Sigma_{t11}^{-1} \Sigma_{t21}$$

$$K_t = \Sigma_{t11}^{-1} - \Sigma_{t11}^{-1} A_t^{(1)} \left( A_t^{(1)T} \Sigma_{t11}^{-1} A_t^{(1)} \right)^{-1} A_t^{(1)T} \Sigma_{t11}^{-1}$$

$$M_t = \Sigma_{t21} K_t + A_t^{(2)} \left( A_t^{(1)T} \Sigma_{t11}^{-1} A_t^{(1)} \right)^{-1} A_t^{(1)T} \Sigma_{t11}^{-1}$$

$$G_t = \Sigma_{t22.1} + \left( A_t^{(2)} - \Sigma_{t21} \Sigma_{t11}^{-1} A_t^{(1)} \right) \\ \times \left( A_t^{(1)T} \Sigma_{t11}^{-1} A_t^{(1)} \right)^{-1} \left( A_t^{(2)} - \Sigma_{t21} \Sigma_{t11}^{-1} A_t^{(1)} \right)^T$$

Our objective is to use all the data available at a given time in predicting

$$\xi_t \equiv \xi_t(C_t^{(1)}, C_t^{(2)}) = P_1 C_t^{(1)} + P_2 C_t^{(2)}$$

where  $P_1$  and  $P_2$  are known matrices. In particular, we may be interested in predicting retrospectively the population totals for the small areas at the previous time points based on all the available sampled data at a given time. To this end, we derive the predictive distribution of  $C_t^{(2)}$  given  $C_t^{(1)} = c_t^{(1)}$  in the following theorem. A proof of this theorem is given in Appendix.

Theorem 4.2.1 Consider the model in 4.2.1 or 4.2.2. Assume that  $\alpha, R, \lambda_1 R, \dots, \lambda_s R$  are independently distributed with

$$\alpha \sim \text{Uniform}(R^p)$$

$$R \sim \text{Gamma}(\tfrac{1}{2}a_0, \tfrac{1}{2}h_0); a_0 \geq 0, h_0 \geq 0$$

$$\lambda_l R \sim \text{Gamma}(\tfrac{1}{2}a_l, \tfrac{1}{2}h_l); l = 1, \dots, s$$

with  $a_l > 0, h_l \geq 0, (l = 1, \dots, s)$ . Assume that  $n_t + \sum_{l=0}^s h_l - p > 2$ . Then conditional on  $\Lambda = \lambda$ , and  $C_t^{(1)} = c_t^{(1)}$ ,  $C_t^{(2)}$  is distributed as multivariate  $t$  with degrees of freedom  $n_t + \sum_{l=0}^s h_l - p$ , location parameter  $M_t c_t^{(1)}$  and scale parameter

$$\Omega_t = \left( n_t + \sum_{l=0}^s h_l - p \right)^{-1} \left( a_0 + \sum_{l=1}^s a_l \lambda_l + c_t^{(1)T} K_t c_t^{(1)} \right) G_t \quad (4.2.3)$$

Also the conditional distribution of  $\Lambda$  given  $C_t^{(1)} = c_t^{(1)}$  has the pdf

$$f(\lambda | c_t^{(1)}) \propto |\Sigma_{t11}|^{-\frac{1}{2}} \left| A_t^{(1)T} \Sigma_{t11}^{-1} A_t^{(1)} \right|^{-\frac{1}{2}} \left[ \prod_{l=1}^s \lambda_l^{\frac{1}{2} h_l - 1} \right] \\ \times \left( a_0 + \sum_{l=1}^s a_l \lambda_l + c_t^{(1)T} K_t c_t^{(1)} \right)^{-\frac{1}{2} (n_t + \sum_{l=0}^s h_l - p)} \quad (4.2.4)$$

Now the conditional expectation and variance of  $C_t^{(2)}$  given  $C_t^{(1)} = c_t^{(1)}$  can be derived as

$$E(C_t^{(2)} | c_t^{(1)}) = E(M_t | c_t^{(1)}) c_t^{(1)} \quad (4.2.5)$$

$$V(C_t^{(2)} | c_t^{(1)}) = V(E(C_t^{(2)} | \Lambda, c_t^{(1)}) | c_t^{(1)}) + E(V(C_t^{(2)} | \Lambda, c_t^{(1)}) | c_t^{(1)}) \\ = V(M_t c_t^{(1)} | c_t^{(1)}) + \left( n_t + \sum_{l=0}^s h_l - p - 2 \right)^{-1} \\ \times E \left[ \left\{ a_0 + \sum_{l=1}^s a_l \lambda_l + c_t^{(1)T} K_t c_t^{(1)} \right\} G_t | c_t^{(1)} \right] \quad (4.2.6)$$

From (4.2.5) and (4.2.6), for  $t \geq 1$  we can find the posterior mean and variance of  $\xi_t \equiv \xi_t(C_t^{(1)}, C_t^{(2)}) = P_1 C_t^{(1)} + P_2 C_t^{(2)}$  where  $P_1$  and  $P_2$  are known matrices. The Bayes estimate of  $\xi_t(C_t^{(1)}, C_t^{(2)})$  under any quadratic loss is its posterior mean and is given by

$$e_t^B(c_t^{(1)}) = [P_1 + P_2 E(M_t | c_t^{(1)})] c_t^{(1)}$$

Also,

$$V[\xi_t(C_t^{(1)}, C_t^{(2)}) | C_t^{(1)}] = P_2 V(C_t^{(2)} | c_t^{(1)}) P_2^T$$

Note that when  $P_1 = \oplus_{i=1}^m P_{1i}$  and  $P_2 = \oplus_{i=1}^m P_{2i}$

where

$$P_{1i} = [\emptyset, \emptyset, \dots, 1_{n_{ij}}^T, \dots, \emptyset]$$



and

$$\mathbf{P}_{2i} = [\emptyset, \emptyset, \dots, \mathbf{1}_{N_{ij}-n_{ij}}^T, \dots, \emptyset]$$

$\xi_t(C_t^{(1)}, C_t^{(2)})$  reduces to the vector of finite population totals for the  $m$  small areas at time  $j$ , ( $j = 1, 2, \dots, t$ ) and whereas for the choice  $\mathbf{P}_1 = \oplus_{i=1}^m \mathbf{P}_{1i}/N_{ij}$  and  $\mathbf{P}_2 = \oplus_{i=1}^m \mathbf{P}_{2i}/N_{ij}$ ,  $\xi_t(C_t^{(1)}, C_t^{(2)})$  reduces to the vector of finite population means for the  $m$  small areas at time  $j$ , ( $j = 1, 2, \dots, t$ ).

Also for  $t \geq 2$ , when  $\mathbf{P}_1 = \oplus_{i=1}^m \mathbf{P}_{1i}$  and  $\mathbf{P}_2 = \oplus_{i=1}^m \mathbf{P}_{2i}$  where

$$\mathbf{P}_{1i} = \left[ \emptyset, \emptyset, \dots, \emptyset, -\frac{1}{N_{i(t-1)}} \mathbf{1}_{n_{i(t-1)}}^T, \frac{1}{N_{it}} \mathbf{1}_{n_{it}}^T \right]$$

and

$$\mathbf{P}_{2i} = \left[ \emptyset, \emptyset, \dots, \emptyset, -\frac{1}{N_{i(t-1)}} \mathbf{1}_{N_{i(t-1)}-n_{i(t-1)}}^T, \frac{1}{N_{it}} \mathbf{1}_{N_{it}-n_{it}}^T \right]$$

$\xi_t(C_t^{(1)}, C_t^{(2)})$  reduces to the vector of change in mean at time  $t$  from previous time point, for the  $m$  small areas.

However, in order to find the conditional distribution of  $C_t^{(2)}$  given  $C_t^{(1)} = \mathbf{c}_t^{(1)}$  analytically, we need to find the posterior density  $f(\boldsymbol{\lambda} | \mathbf{c}_t^{(1)})$  as given in (4.2.4). Even when  $\boldsymbol{\lambda}$  is a small-dimensional vector, it is not easy to find this density through the use of sophisticated numerical or analytic approximation techniques. In contrast, this posterior density as well as the conditional expectation and variance in (4.2.5) and (4.2.6) can be found with considerable ease in a more general case when  $\psi_j$ , ( $j = 1, \dots, t$ ) and  $\mathbf{W}$  are unknown, by employing the Gibbs sampling algorithm. It is assumed that  $V_{ij}$  is known. In such a case, we rewrite the model as:

$$\text{I. } \left( \begin{array}{c} \mathbf{Y}_{ij}^{(1)} \\ \mathbf{Y}_{ij}^{(2)} \end{array} \right) \left| \left( \begin{array}{c} \boldsymbol{\theta}_{ij}^{(1)} \\ \boldsymbol{\theta}_{ij}^{(2)} \end{array} \right), R \stackrel{\text{ind}}{\sim} N \left( \left( \begin{array}{c} \boldsymbol{\theta}_{ij}^{(1)} \\ \boldsymbol{\theta}_{ij}^{(2)} \end{array} \right), r^{-1} \mathbf{V}_{ij} \right);$$

$$\text{where } \mathbf{V}_{ij} = \begin{bmatrix} \mathbf{V}_{ij11} & \mathbf{V}_{ij12} \\ \mathbf{V}_{ij21} & \mathbf{V}_{ij22} \end{bmatrix},$$

$$(i = 1, \dots, m; j = 1, \dots, t)$$

$$\text{II. } \left( \begin{array}{c} \boldsymbol{\theta}_{ij}^{(1)} \\ \boldsymbol{\theta}_{ij}^{(2)} \end{array} \right) \left| \boldsymbol{\alpha}, \mathbf{b}_j, R \stackrel{\text{ind}}{\sim} N \left( \left( \begin{array}{c} \mathbf{X}_{ij}^{(1)} \\ \mathbf{X}_{ij}^{(2)} \end{array} \right) \boldsymbol{\alpha} + \left( \begin{array}{c} \mathbf{Z}_{ij}^{(1)} \\ \mathbf{Z}_{ij}^{(2)} \end{array} \right) \mathbf{b}_j, r^{-1} \boldsymbol{\psi}_j \right);$$

$$(i = 1, \dots, m; j = 1, \dots, t)$$

$$\text{III. } \mathbf{b}_j | \mathbf{b}_{j-1}, \mathbf{W}, R \stackrel{\text{ind}}{\sim} N(\mathbf{H}_j \mathbf{b}_{j-1}, r^{-1} \mathbf{W}); (j = 1, \dots, t)$$

We assume that  $\boldsymbol{\alpha}, R, \boldsymbol{\psi}_j, \dots, \boldsymbol{\psi}_t$ , and  $\mathbf{W}$  are mutually independent with

$$\boldsymbol{\alpha} \sim \text{Uniform}(R^p),$$

$$R \sim \text{Gamma}(\tfrac{1}{2}a_0, \tfrac{1}{2}h_0) \quad a_0 \geq 0, \quad h_0 \geq 0,$$

$$\boldsymbol{\psi}_j \sim \text{inverse Wishart}(\mathbf{0}, \mathbf{k}_j),$$

$$\mathbf{W} \sim \text{inverse Wishart}(\mathbf{0}, \mathbf{k}_0).$$

The Gibbs sampling algorithm requires samples from the following complete conditional distributions:

- $\boldsymbol{\alpha} | \mathbf{c}_t^{(1)}, \mathbf{c}_t^{(2)}, \boldsymbol{\theta}, R, \mathbf{b}_1, \dots, \mathbf{b}_t, \boldsymbol{\psi}_1, \dots, \boldsymbol{\psi}_t, \mathbf{W}$
- $\mathbf{b}_j | \mathbf{c}_t^{(1)}, \mathbf{c}_t^{(2)}, \boldsymbol{\theta}, R, \boldsymbol{\alpha}, \boldsymbol{\psi}_1, \dots, \boldsymbol{\psi}_t, \mathbf{b}_1, \dots, \mathbf{b}_{j-1}, \mathbf{b}_{j+1}, \dots, \mathbf{b}_t, \mathbf{W}, j = 1, \dots, t$
- $\boldsymbol{\psi}_1, \dots, \boldsymbol{\psi}_t | \mathbf{c}_t^{(1)}, \mathbf{c}_t^{(2)}, \boldsymbol{\theta}, R, \boldsymbol{\alpha}, \mathbf{b}_1, \dots, \mathbf{b}_t, \mathbf{W}$
- $\mathbf{W} | \mathbf{c}_t^{(1)}, \mathbf{c}_t^{(2)}, \boldsymbol{\theta}, R, \boldsymbol{\alpha}, \mathbf{b}_1, \dots, \mathbf{b}_t, \boldsymbol{\psi}_1, \dots, \boldsymbol{\psi}_t$
- $\mathbf{Y}_{ij}^{(2)} | \mathbf{c}_t^{(1)}, \boldsymbol{\theta}, R, \mathbf{b}_1, \dots, \mathbf{b}_t, \boldsymbol{\psi}_1, \dots, \boldsymbol{\psi}_t, \mathbf{W}, (i = 1, \dots, m; j = 1, \dots, t)$

The hierarchical structure of the model and the conjugacy of the priors lead to the following complete conditional distribution:

$$\begin{aligned}
(i) \quad & \alpha | c_t^{(1)}, c_t^{(2)}, \theta, r, b_1, \dots, b_t, \psi_1, \dots, \psi_t, W \\
& \sim \alpha | c_t^{(1)}, \theta, r, b_1, \dots, b_t, \psi_1, \dots, \psi_t \\
& \sim N_p \left[ \left( \sum_{i=1}^m \sum_{j=1}^t X_{ij}^T \psi_j^{-1} X_{ij} \right)^{-1} \right. \\
& \quad \left. \sum_{i=1}^m \sum_{j=1}^t X_{ij}^T \psi_j^{-1} (\theta_{ij} - Z_{ij} b_j), \right. \\
& \quad \left. \left( \sum_{i=1}^m \sum_{j=1}^t X_{ij}^T \psi_j^{-1} X_{ij} \right)^{-1} \right]
\end{aligned}$$

(iia) For  $(j = 1, \dots, t-1)$  :

$$\begin{aligned}
& b_j | c_t^{(1)}, c_t^{(2)}, \theta, r, \alpha, b_1, \dots, b_{j-1}, b_{j+1}, \dots, b_t, \psi_1, \dots, \psi_t, W \\
& \sim b_j | \theta, r, \alpha, b_1, \dots, b_{j-1}, b_{j+1}, \dots, b_t, \psi_1, \dots, \psi_t, W \\
& \sim N_q \left[ \left( \sum_{i=1}^m Z_{ij}^T \psi_j^{-1} Z_{ij} + H_{j+1}^T W^{-1} H_{j+1} + W^{-1} \right)^{-1} \right. \\
& \quad \times \left\{ \sum_{i=1}^m Z_{ij}^T \psi_j^{-1} (\theta_{ij} - X_{ij} \alpha) \right. \\
& \quad \left. \left. + H_{j+1}^T W^{-1} b_{j+1} + W^{-1} H_j b_{j-1} \right\}, \right. \\
& \quad \left. \left( \sum_{i=1}^m Z_{ij}^T \psi_j^{-1} Z_{ij} + H_{j+1}^T W^{-1} H_{j+1} + W^{-1} \right)^{-1} \right]
\end{aligned}$$

$$\begin{aligned}
(iib) \quad & b_t | c_t^{(1)}, c_t^{(2)}, \theta, r, \alpha, b_1, \dots, b_{t-1}, \psi_1, \dots, \psi_t, W \\
& \sim b_t | \theta, r, \alpha, b_1, \dots, b_{t-1}, \psi_1, \dots, \psi_t, W \\
& \sim N_q \left[ \left( \sum_{i=1}^m Z_{it}^T \psi_t^{-1} Z_{it} + W^{-1} \right)^{-1} \right. \\
& \quad \times \left\{ \sum_{i=1}^m Z_{it}^T \psi_t^{-1} (\theta_{it} - X_{it} \alpha) + W^{-1} H_t b_{t-1} \right\}, \\
& \quad \left. \left( \sum_{i=1}^m Z_{it}^T \psi_t^{-1} Z_{it} + W^{-1} \right)^{-1} \right]
\end{aligned}$$

$$(iii) \quad \psi_1, \dots, \psi_t | \mathbf{c}_t^{(1)}, \mathbf{c}_t^{(2)}, r, \boldsymbol{\alpha}, \boldsymbol{\theta}, \mathbf{b}_1, \dots, \mathbf{b}_t, \mathbf{W}$$

$$\sim \psi_1, \dots, \psi_t | r, \boldsymbol{\alpha}, \boldsymbol{\theta}, \mathbf{b}_1, \dots, \mathbf{b}_t$$

$$\stackrel{ind}{\sim} \text{Inverse Wishart} \left[ r \sum_{i=1}^m (\boldsymbol{\theta}_{ij} - \mathbf{X}_{ij}\boldsymbol{\alpha} - \mathbf{Z}_{ij}\mathbf{b}_j) \right. \\ \left. \times (\boldsymbol{\theta}_{ij} - \mathbf{X}_{ij}\boldsymbol{\alpha} - \mathbf{Z}_{ij}\mathbf{b}_j)^T, m + k_j \right]$$

$$(iv) \quad \mathbf{W} | \mathbf{c}_t^{(1)}, \mathbf{c}_t^{(2)}, \boldsymbol{\alpha}, \boldsymbol{\theta}, \mathbf{b}_1, \dots, \mathbf{b}_t, \psi_1, \dots, \psi_t$$

$$\sim \mathbf{W} | r, \mathbf{b}_1, \dots, \mathbf{b}_t$$

$$\sim \text{Inverse Wishart} \left( r \sum_{j=1}^t (\mathbf{b}_j - \mathbf{H}_j \mathbf{b}_{j-1}) (\mathbf{b}_j - \mathbf{H}_j \mathbf{b}_{j-1})^T, t + k_0 \right)$$

$$(v) \quad \boldsymbol{\theta}_{ij} | \mathbf{c}_t^{(1)}, \mathbf{c}_t^{(2)}, \boldsymbol{\alpha}, \mathbf{b}_1, \dots, \mathbf{b}_t, \psi_1, \dots, \psi_t, \mathbf{W}$$

$$\sim \boldsymbol{\theta}_{ij} | \mathbf{c}_t^{(1)}, \mathbf{c}_t^{(2)}, \boldsymbol{\alpha}, \mathbf{b}_1, \dots, \mathbf{b}_t, \psi_1, \dots, \psi_t, \mathbf{W}$$

$$\sim N \left[ \left( \mathbf{V}_{ij}^{-1} + \boldsymbol{\psi}_j^{-1} \right)^{-1} \left( \mathbf{V}_{ij}^{-1} \mathbf{y}_{ij} + \boldsymbol{\psi}_j^{-1} (\mathbf{X}_{ij}\boldsymbol{\alpha} + \mathbf{Z}_{ij}\mathbf{b}_j) \right), \right. \\ \left. \left( \mathbf{V}_{ij}^{-1} + \boldsymbol{\psi}_j^{-1} \right)^{-1} \right]$$

$$(vi) \quad \mathbf{y}_{ij}^{(2)} | \mathbf{c}_t^{(1)}, \boldsymbol{\theta}, r, \boldsymbol{\alpha}, \mathbf{b}_1, \dots, \mathbf{b}_t, \psi_1, \dots, \psi_t, \mathbf{W}$$

$$\sim N_{N_{ij}-n_{ij}} \left[ \boldsymbol{\theta}_{ij}^{(2)} + \mathbf{V}_{ij21} \mathbf{V}_{ij11}^{-1} \left( \mathbf{y}_{ij}^{(1)} - \boldsymbol{\theta}_{ij}^{(1)} \right), \mathbf{V}_{ij21} \mathbf{V}_{ij11}^{-1} \right]$$

### 4.3 The Hierarchical Bayes Predictor in a Special Case

This section considers a special case of the HB model in Section 4.2. We assume that the vector of the ratios of variance components,  $\boldsymbol{\lambda}$  is known. Our target is again to find an optimal predictor for

$$\boldsymbol{\xi}_t \equiv \boldsymbol{\xi}_t(\mathbf{C}_t^{(1)}, \mathbf{C}_t^{(2)}) = \mathbf{P}_1 \mathbf{C}_t^{(1)} + \mathbf{P}_2 \mathbf{C}_t^{(2)}$$

for some appropriate  $\mathbf{P}_1$  and  $\mathbf{P}_2$ , based on all the available data at a given time. To this end, it suffices to find the predictive distribution of  $\mathbf{C}_t^{(2)}$  given  $\mathbf{C}_t^{(1)} = \mathbf{c}_t^{(1)}$ . This is done in the following theorem.

Theorem 4.3.1 Consider the model given in 4.2.1 or 4.2.2 with  $\lambda$  known. Assume that  $\alpha$  and  $R$  are independently distributed with

$$\alpha \sim \text{Uniform}(R^p)$$

$$R \sim \text{Gamma}(\tfrac{1}{2}a_0, \tfrac{1}{2}h_0); a_0 \geq 0, h_0 \geq 0$$

Also, assume that  $n_t + h_0 - p > 2$ . Then conditional on  $C_t^{(1)} = \mathbf{c}_t^{(1)}$ ,  $C_t^{(2)}$  is distributed as multivariate  $t$  with degrees of freedom  $n_t + h_0 - p$ , location parameter  $M_t \mathbf{c}_t^{(1)}$  and scale parameter

$$\Omega_t = (n_t + h_0 - p)^{-1} \left( a_0 + \mathbf{c}_t^{(1)T} \mathbf{K}_t \mathbf{c}_t^{(1)} \right) \mathbf{G}_t \quad (4.3.7)$$

Proof of this theorem is similar to that of Theorem 4.2.1 and is omitted.

In this case, it is possible to obtain the Bayes estimate of  $\xi_t \equiv \xi_t(C_t^{(1)}, C_t^{(2)})$  in a closed form and is given by

$$\mathbf{e}_t^{B*}(\mathbf{c}_t^{(1)}) = \left[ \mathbf{P}_1 + \mathbf{P}_2 E(\mathbf{M}_t | \mathbf{c}_t^{(1)}) \right] \mathbf{c}_t^{(1)} \quad (4.3.8)$$

Remark 4.3.1 We may note that the predictor  $\mathbf{e}_t^{B*}(\mathbf{c}_t^{(1)})$  given in (4.3.8) can be obtained without using any numerical integration techniques or the Gibbs sampling algorithm.

Remark 4.3.2 The predictor  $\mathbf{e}_t^{B*}(\mathbf{c}_t^{(1)})$  does not depend on the choice of the prior (proper) distribution of  $R$ . In this sense, the predictor  $\mathbf{e}_t^{B*}(\mathbf{c}_t^{(1)})$  is robust against the choice of priors for  $R$ . To see this,



$$\begin{aligned}
E(C_t^{(2)}|C_t^{(1)}) &= E[E\{E(C_t^{(2)}|\alpha, R, C_t^{(1)})|R, C_t^{(1)}\}|C_t^{(1)}] \\
&= E[E\{A_t^{(2)}\alpha + \Sigma_{t21}\Sigma_{t11}^{-1}(C_t^{(1)} - A_t^{(1)}\alpha)|R, C_t^{(1)}\}|C_t^{(1)}] \\
&= \Sigma_{t21}\Sigma_{t11}^{-1}C_t^{(1)} + (A_t^{(2)} - \Sigma_{t21}\Sigma_{t11}^{-1}A_t^{(1)}) \\
&\quad \times E[E\{\alpha|R, C_t^{(1)}\}|C_t^{(1)}] \\
&= \Sigma_{t21}\Sigma_{t11}^{-1}C_t^{(1)} + (A_t^{(2)} - \Sigma_{t21}\Sigma_{t11}^{-1}A_t^{(1)}) \\
&\quad \times (A_t^{(1)T}\Sigma_{t11}^{-1}A_t^{(1)})^{-1}A_t^{(1)T}\Sigma_{t11}^{-1}C_t^{(1)} \\
&= M_t C_t^{(1)},
\end{aligned}$$

It is assumed that all the expectations appearing above exist.

**Remark 4.3.3** The predictor  $e_t^{B*}(c_t^{(1)})$  of  $\xi_t \equiv \xi_t(C_t^{(1)}, C_t^{(2)})$  can also be obtained in an alternative manner. Suppose, first  $\alpha$  known and  $r$  may be known or unknown. In such a case, the best predictor (best linear predictor without the normality assumption) of  $\xi_t(C_t^{(1)}, C_t^{(2)})$  in the sense of having the smallest mean squared error matrix is given by

$$\begin{aligned}
E_{\alpha, R}[\xi_t(C_t^{(1)}, C_t^{(2)})|C_t^{(1)}] &= A_t^{(1)}C_t^{(1)} + P_2[\Sigma_{t21}\Sigma_{t11}^{-1}C_t^{(1)} \\
&\quad + (A_t^{(2)} - \Sigma_{t21}\Sigma_{t11}^{-1}A_t^{(1)})\alpha] \text{ a.e. } (C_t^{(1)})
\end{aligned}$$

Now, when  $\alpha$  is unknown, then one replaces  $\alpha$  by its UMVUE

$(A_t^{(1)T}\Sigma_{t11}^{-1}A_t^{(1)})^{-1}A_t^{(1)T}\Sigma_{t11}^{-1}C_t^{(1)}$ . The resulting predictor of  $\xi_t \equiv \xi_t(C_t^{(1)}, C_t^{(2)})$  turns out to be  $e_t^{B*}(c_t^{(1)})$ . In this sense,  $e_t^{B*}(c_t^{(1)})$  is also an EB predictor of  $\xi_t(C_t^{(1)}, C_t^{(2)})$ .

#### 4.4 Best Unbiased Prediction in Small Area Estimation

In this section, we prove the optimality of the predictor  $e_t^{B*}$  within the class of all unbiased predictors of  $\xi_t(C_t^{(1)}, C_t^{(2)})$ . We consider the model given in 4.2.1 or

4.2.2 with  $\lambda$  known, but  $\alpha$  and  $r = (\sigma^2)^{-1}$  unknown. We do not assume any prior distributions for  $\alpha$  and  $r$  and treat them as unknown parameters.

Write  $\Phi = (\alpha^T, r)^T$ .

**Definition 4.4.1.** A predictor  $T(C_t^{(1)})$  is said to be a BUP of  $\xi_t(C_t^{(1)}, C_t^{(2)})$  if  $E_\Phi [T(C_t^{(1)}) - \xi_t(C_t^{(1)}, C_t^{(2)})] = 0$  for all  $\Phi$  and for every predictor  $\delta(C_t^{(1)})$  of  $\xi_t(C_t^{(1)}, C_t^{(2)})$  satisfying

$$E_\Phi [\delta(C_t^{(1)}) - \xi_t(C_t^{(1)}, C_t^{(2)})] = 0 \text{ for all } \Phi \quad (4.4.9)$$

$$V_\Phi [\delta(C_t^{(1)}) - \xi_t(C_t^{(1)}, C_t^{(2)})] - V_\Phi [T(C_t^{(1)}) - \xi_t(C_t^{(1)}, C_t^{(2)})] \quad (4.4.10)$$

is non-negative definite (n.n.d) for all  $\Phi$  provided the quantities are finite.

The BUP property of  $e_t^{B*}$  is proved in the following theorem.

**Theorem 4.4.1** Consider the model given in 4.2.1 or 4.2.2 with  $\alpha$  and  $r = (\sigma^2)^{-1}$  treated as unknown parameters but  $\lambda$  assumed to be known.

$e_t^{B*}(C_t^{(1)}) = (P_1 + P_2 M_t) C_t^{(1)}$  is the BUP of  $\xi_t(C_t^{(1)}, C_t^{(2)})$ .

The proof of the theorem is facilitated by the following a general lemma which is a simple extension of Lemma 3.3.1 of Datta (1990). The lemma concerns unbiased prediction of a general  $g(C_t)$  ( $u \times 1$ ) based on  $C_t^{(1)}$ , where  $g(C_t)$  is not necessarily equal to  $\xi_t(C_t^{(1)}, C_t^{(2)})$  or linear in  $C_t$ . It is assumed that a finite second moment exists for each component  $g_i(C_t), i = 1, \dots, u$ .

**Lemma 4.4.1** A predictor  $T(C_t^{(1)}) \in U_g$  is BUP for  $g(C_t)$  if and only if

$$Cov_\Phi [T(C_t^{(1)}) - g(C_t^{(1)}), m(C_t^{(1)})] = 0 \quad (4.4.11)$$

for all  $g(C_t)$  and for every  $m \in U_0$ ,  $U_g \equiv$  Class of all unbiased predictors  $\delta(C_t^{(1)})$  of  $g(C_t)$  with each component of  $\delta(C_t^{(1)})$  having a finite second moment.  $U_0$  is the

class of all real valued statistics (i.e., functions of  $C_t^{(1)}$ ) with finite second moments having zero expectations identically in  $\Phi$ .

Proof of Lemma 4.4.1 Let  $T(C_t^{(1)}) = (T_1(C_t^{(1)}), \dots, T_u(C_t^{(1)}))^T$ .

If  $\delta(C_t^{(1)}) = (\delta_1(C_t^{(1)}), \dots, \delta_u(C_t^{(1)}))^T$  is another predictor in  $U_g$ , then

$$\begin{aligned} V_{\Phi} [\delta(C_t^{(1)}) - g(C_t)] = \\ V_{\Phi} [T(C_t^{(1)}) - g(C_t)] + V_{\Phi} [\delta(C_t^{(1)}) - T(C_t)] \\ + Cov_{\Phi} [T(C_t^{(1)}) - g(C_t), \delta(C_t^{(1)}) - T(C_t)] \\ + Cov_{\Phi} [\delta(C_t^{(1)}) - T(C_t), T(C_t^{(1)}) - g(C_t)] \end{aligned} \quad (4.4.12)$$

Now  $T_i(C_t^{(1)}) - \delta_i(C_t^{(1)}) \in U_0$  for every  $i = 1, \dots, u$ .

Hence, using the condition of the lemma, for  $1 \leq i \leq u$ ,

$$Cov_{\Phi} [T(C_t^{(1)}) - g(C_t), T_i(C_t^{(1)}) - \delta_i(C_t^{(1)})] = 0. \quad (4.4.13)$$

From (4.4.12) and (4.4.13) it follows that

$$\begin{aligned} V_{\Phi} [\delta(C_t^{(1)}) - g(C_t)] = V_{\Phi} [T(C_t^{(1)}) - g(C_t)] \\ + V_{\Phi} [\delta(C_t^{(1)}) - T(C_t)] \end{aligned} \quad (4.4.14)$$

for all  $\Phi$ . Hence  $T(C_t^{(1)})$  is BUP for  $g(C_t)$ .

Only if. Given that  $T(C_t^{(1)})$  is BUP, we will show that the condition (4.4.11) is true.

First we will show that  $T_i(C_t^{(1)})$  is BUP for  $g_i(C_t)$  for every  $i = 1, \dots, u$ . Let  $U_i(C_t^{(1)})$  be any unbiased predictor for  $g_i(C_t)$ . Then  $\delta^*(C_t^{(1)})$ , a  $u$ -component vector with  $i$ th component equal to  $U_i(C_t^{(1)})$  belongs to  $U_g$ .

Then

$$V_{\Phi} [\delta^*(C_t^{(1)}) - g(C_t)] - V_{\Phi} [T(C_t^{(1)}) - g(C_t)]$$

is n.n.d. by hypothesis.

So we have

$$V_{\Phi} [U_i(C_t^{(1)}) - g_i(C_t)] - V_{\Phi} [T_i(C_t^{(1)}) - g_i(C_t)] \geq 0$$

and consequently  $T_i(C_t^{(1)})$  is BUP for  $g_i(C_t)$ .

Now following the usual Lehmann-Scheffe (1950) technique (also Rao, 1952), for any  $m \in U_0$

$$\text{Cov}_{\Phi} [T_i(C_t^{(1)}) - g_i(C_t), m(C_t^{(1)})] = 0 \quad (4.4.15)$$

for all  $1 \leq i \leq u$ . Hence, (4.4.11) holds, and the proof of the lemma is complete.

**Remark 4.4.1.** It follows from the proof of the above lemma that the BUP of  $g(C_t)$  is unique with probability one.

Proof of Theorem 4.4.1 In view of Lemma 4.4.1, it suffices to show that for every  $m(C_t^{(1)}) \in U_0$ ,

$$\text{Cov}_{\Phi} [e_t^{B*}(C_t^{(1)}) - \xi_t(C_t^{(1)}, C_t^{(2)}), m(C_t^{(1)})] = 0 \text{ for all } \Phi.$$

$$\text{that is } E_{\Phi} [P_1(M_t C_t^{(1)} - C_t^{(2)}) m(C_t^{(1)})] = 0 \text{ for all } \Phi.$$

Since, under the model,

$$\begin{aligned} M_t C_t^{(1)} - E_{\Phi}(C_t^{(2)} | C_t^{(1)}) &= (A_t^{(2)} - \Sigma_{t21} \Sigma_{t11}^{-1} A_t^{(1)}) \\ &\quad \times \left[ (A_t^{(1)T} \Sigma_{t11}^{-1} A_t^{(1)})^{-1} A_t^{(1)T} \Sigma_{t11}^{-1} c_t^{(1)} - \alpha \right], \end{aligned}$$

using  $E_{\Phi}[m(C_t^{(1)})] = 0$ , it suffices to show that

$$E_{\Phi} \left[ (A_t^{(1)T} \Sigma_{t11}^{-1} c_t^{(1)}) m(C_t^{(1)}) \right] = 0 \quad (4.4.16)$$

for all  $\Phi$ . Since  $E_{\Phi}[m(C_t^{(1)})] = 0$ ,

$$\Leftrightarrow \int m(C_t^{(1)}) \exp \left[ -\frac{1}{2} r (c_t^{(1)} - A_t^{(1)} \alpha)^T \Sigma_{t11}^{-1} (c_t^{(1)} - A_t^{(1)} \alpha) \right] d c_t^{(1)} = 0, \quad (4.4.17)$$

differentiating both sides of this equation w.r.t  $\alpha$ , one gets (see p 318 of Rao, 1973)

$$\begin{aligned} & \int A_t^{(1)T} \Sigma_{t11}^{-1} (c_t^{(1)} - A_t^{(1)} \alpha) m(C_t^{(1)}) \\ & \quad \times \exp \left[ \frac{1}{2} r (c_t^{(1)} - A_t^{(1)} \alpha)^T \Sigma_{t11}^{-1} (c_t^{(1)} - A_t^{(1)} \alpha) \right] d c_t^{(1)} \\ & = 0. \end{aligned} \quad (4.4.18)$$

Using  $E_{\Phi}[m(C_t^{(1)})] = 0$  again, (4.4.16) follows from (4.4.18). The proof of Theorem is complete.

Remark 4.4.2. We can also prove Equation (4.4.16) in the following way. Note that since  $C_t^{(1)} \sim N(A_t^{(1)} \alpha, r^{-1} \Sigma_{t11})$ ,  $(A_t^{(1)T} \Sigma_{t11}^{-1} C_t^{(1)}, C_t^{(1)T} K_t C_t^{(1)})$  is complete sufficient  $\Phi$ . Hence  $A_t^{(1)T} \Sigma_{t11}^{-1} C_t^{(1)}$  must have 0 covariance with every zero estimator  $m(C_t^{(1)})$ , i.e.,  $E_{\Phi}[(A_t^{(1)T} \Sigma_{t11}^{-1} C_t^{(1)}) m(C_t^{(1)})] = 0$ .

Next we generalize Theorem 4.4.1 for certain non-normal general distributions. Suppose that

$$e^* = (\omega^T, \eta^T)^T$$

$$\Delta = \text{Block Diagonal} (B_t(I_t \otimes W)B_t^T, I_m \otimes \Psi_t + V_t)$$

Assume that given  $R = r$ ,  $e^* \sim N(0, r^{-1} \Delta)$ , while marginally the distribution function of  $R$  is an arbitrary member of the family  $\mathcal{F} = \{F : F \text{ is absolutely continuous with pdf } f(r) = 0 \text{ for } r \leq 0\}$ . Let  $\mathcal{F}^*$  denote a subfamily of  $\mathcal{F}$  such that each component of  $e_t^{B^*}(C_t^{(1)})$  and  $\xi_t(C_t^{(1)}, C_t^{(2)})$  has finite second moment under the model 4.2.1 and the joint distribution of  $e^*$  and  $R$ . In such a set up, we now prove the BUP property of the predictor  $e_t^{B^*}(C_t^{(1)})$  in the following theorem which is a simple extension of the Theorem 3.3.2. of Datta (1990).



Theorem 4.4.2  $e_t^{B*}(C_t^{(1)})$  is BUP of  $\xi_t(C_t^{(1)}, C_t^{(2)})$  under the model in 4.2.1,  $e^*|R = r \sim N(0, r^{-1}\Delta)$  and  $R$  has a df from  $\mathcal{F}^*$ .

Proof of Theorem 4.4.2 . Using lemma 4.4.1 and following the proof of Theorem 4.4.1, it suffices to show that

$$E_{\alpha, F} \left[ (A_t^{(1)})^T \Sigma_{t11}^{-1} C_t^{(1)} m(C_t^{(1)}) \right] = 0 \quad (4.4.19)$$

for all  $\alpha \in R^p$  and for all  $F \in \mathcal{F}^*$ , where

$$E_{\alpha, F} [m(C_t^{(1)})] = 0 \quad (4.4.20)$$

and  $E_{\alpha, F} [m^2(C_t^{(1)})] < \infty$  for all  $\alpha$  and for all  $F \in \mathcal{F}^*$ . Consider the subfamily  $\mathcal{H} = \{R \sim \text{gamma}(\frac{1}{2}c, \frac{1}{2}d) : c > 0, d > 2\}$  of  $\mathcal{F}^*$ .

Since (4.4.20) holds for this subfamily  $\mathcal{H}$ ,  $E_{\alpha, F} [m(C_t^{(1)})] = 0$  for all  $\alpha$  and for all  $F \in \mathcal{H}$  gives

$$\begin{aligned} & \int_0^\infty \exp(-\frac{1}{2}cr) \int r^{\frac{1}{2}(n_t+d)-1} \exp \left[ -\frac{1}{2}r(\mathbf{c}_t^{(1)} - A_t^{(1)}\alpha)^T \Sigma_{t11}^{-1}(\mathbf{c}_t^{(1)} - A_t^{(1)}\alpha) \right] \\ & \times m(\mathbf{c}_t^{(1)}) d\mathbf{c}_t^{(1)} dr = 0 \end{aligned} \quad (4.4.21)$$

for all  $\alpha$ ,  $c > 0$  and  $d > 2$ . Now using the uniqueness property of Laplace transforms, it follows from (4.4.21) that

$$\begin{aligned} & \int r^{\frac{1}{2}(n_t+d)-1} \exp \left[ -\frac{1}{2}r(\mathbf{c}_t^{(1)} - A_t^{(1)}\alpha)^T \Sigma_{t11}^{-1}(\mathbf{c}_t^{(1)} - A_t^{(1)}\alpha) \right] \\ & \times m(\mathbf{c}_t^{(1)}) d\mathbf{c}_t^{(1)} = 0 \end{aligned}$$

a.e. Lebesgue for all  $r > 0$  and all  $\alpha$ , i.e.,

$$\begin{aligned} & \int \exp \left[ -\frac{1}{2}r(\mathbf{c}_t^{(1)} - A_t^{(1)}\alpha)^T \Sigma_{t11}^{-1}(\mathbf{c}_t^{(1)} - A_t^{(1)}\alpha) \right] \\ & \times m(\mathbf{c}_t^{(1)}) d\mathbf{c}_t^{(1)} = 0 \end{aligned} \quad (4.4.22)$$

a.e. Lebesgue for all  $r > 0$  and all  $\alpha$ .

Differentiation of both sides of (4.4.22) with respect to  $\alpha$  and some simplifications using (4.4.22) lead to

$$\int (A_t^{(1)T} \Sigma_{t11}^{-1} c_t^{(1)}) m(c_t^{(1)}) \times \exp \left[ -\frac{1}{2} r (c_t^{(1)} - A_t^{(1)} \alpha)^T \Sigma_{t11}^{-1} (c_t^{(1)} - A_t^{(1)} \alpha) \right] d c_t^{(1)} = 0 \quad (4.4.23)$$

a.e. Lebesgue for all  $r > 0$  and all  $\alpha$ .

Multiplying both sides of (4.4.23) by  $r^{\frac{1}{2}n_t}$  and integrating with respect to  $dF(r)$  where  $F \in \mathcal{F}^*$ , one gets (4.4.19).

**Remark 4.4.3.** Since  $\mathcal{F}^*$  does not contain the degenerate distributions of  $R$  on  $(0, \infty)$ , Theorem 4.4.1 does not follow from Theorem 4.4.2.

**Remark 4.4.4.** In Theorem 4.4.2, if we take  $\mathcal{H}$  for  $\mathcal{F}^*$ , we see that the marginal distribution of  $C_t$  is given by the family of distributions

$\{\mathcal{F}(\cdot | N_t, A_t \alpha, (c/d) \Sigma_t, d) : \alpha \in R^p, c > 0, d > 2\}$  and  $e_t^{B*}(c_t^{(1)})$  is BUP for  $\xi_t(C_t^{(1)}, C_t^{(2)})$  for this family where  $\mathcal{F}(\cdot | N_t, A_t \alpha, (c/d) \Sigma_t, d)$  is  $N_t$ -variate  $t$ -distribution with location parameter  $A_t \alpha$ , scale parameter  $(c/d) \Sigma_t$  and degrees of freedom  $d$ . Next we dispense with any distributional assumptions in (4.2.1) and show that  $e_t^{B*}(c_t^{(1)})$  is the BUP of  $\xi_t(C_t^{(1)}, C_t^{(2)})$  within the class of all linear unbiased predictors. A predictor  $\delta(C_t^{(1)})$  is said to be linear if  $\delta(C_t^{(1)})$  has the form  $L_t C_t^{(1)}$  for some known  $u \times n_t$  matrix  $L_t$ . If in addition  $E_{\Phi} [\delta(C_t^{(1)}) - \xi_t(C_t^{(1)}, C_t^{(2)})] = 0$  for all  $\Phi$ , we say that  $\delta(C_t^{(1)})$  is a LUP of  $\xi_t(C_t^{(1)}, C_t^{(2)})$ . We now introduce another definition.

**Definition 4.4.2.** A LUP  $S_t C_t^{(1)}$  of  $\xi_t(C_t^{(1)}, C_t^{(2)})$  is said to be a BLUP if for every LUP  $L_t C_t^{(1)}$  of  $\xi_t(C_t^{(1)}, C_t^{(2)})$ ,  $V_{\Phi} [L_t C_t^{(1)} - \xi_t(C_t^{(1)}, C_t^{(2)})] - V_{\Phi} [S_t C_t^{(1)} - \xi_t(C_t^{(1)}, C_t^{(2)})]$  is n.n.d. for all  $\Phi$ .

We now prove the BLUP property of  $e_t^{B*}(C_t^{(1)})$  in predicting  $\xi_t(C_t^{(1)}, C_t^{(2)})$  in the following theorem.

Theorem 4.4.3 Consider the model in (4.2.1) and assume that  $E_{\Phi}(\omega) = 0$ ,  $E_{\Phi}(\eta) = 0$ ,  $E_{\Phi}(\omega\omega^T) = 0$ ,  $E_{\Phi}(\eta\eta^T) = 0$ ,  $E_{\Phi}(\eta^T\eta) < \infty$  and  $E_{\Phi}(\omega^T\omega) < \infty$ . Then  $e_t^{B*}(C_t^{(1)})$  is the BLUP of  $\xi_t(C_t^{(1)}, C_t^{(2)})$ .

Proof of this theorem requires the following lemma.

Lemma 4.4.2 A LUP  $S_t C_t^{(1)}$  of  $\xi_t(C_t^{(1)}, C_t^{(2)})$  is a BLUP if and only if

$$\text{Cov}_{\Phi} [S_t C_t^{(1)} - \xi_t(C_t^{(1)}, C_t^{(2)}), m^T C_t^{(1)}] = 0$$

for all  $\Phi$  and for every known  $n_t \times 1$  vector  $m$  satisfying  $E_{\Phi}(m^T C_t^{(1)}) = 0$  for all  $\Phi$ .

Proof of this lemma is similar to the proof of Lemma 4.4.1 and is omitted.

Proof of Theorem 4.4.3 If  $E_{\Phi}(m^T C_t^{(1)}) = m^T A_t^{(1)} \alpha = 0$  for all  $\alpha$ ,  $m^T A_t^{(1)} = 0^T$ .

Hence

$$\begin{aligned} & \text{Cov}_{\Phi} [e_t^{B*}(C_t^{(1)}) - \xi_t(C_t^{(1)}, C_t^{(2)}), m^T C_t^{(1)}] \\ &= \text{Cov}_{\Phi} [P_2(M_t C_t^{(1)} - C_t^{(2)}), m^T C_t^{(1)}] \\ &= P_2(M_t \Sigma_{t11} - \Sigma_{t21})m \\ &= P_2(A_t^{(2)} - \Sigma_{t21} \Sigma_{t11}^{-1} A_t^{(1)})(A_t^{(1)T} \Sigma_{t11}^{-1} A_t^{(1)})^{-1} A_t^{(1)T} m \\ &= 0 \end{aligned}$$

for all  $\Phi$ . The last two equalities follow from the definition of  $M_t$  and from the fact that  $m^T A_t^{(1)} = 0^T$ . Applying lemma 4.4.2, the result follows.

Remark 4.4.5. It can be easily proved that BLUP of  $g(C_t)$  is unique with probability one.

#### 4.5 Stochastic Domination

This section examines the stochastic dominance (as in Hwang (1985)) of the HB predictor  $e_t^{B*}(c_t^{(1)})$  within the class of all linear unbiased predictors of  $\xi_t(C_t^{(1)}, C_t^{(2)})$ . We will consider the model of Section 4.2 with  $\lambda$  known. As in the previous Section, we will treat  $\alpha$  and  $R$  as unknown parameters. Also, we write  $e^* = (\omega^T, \eta^T)^T$  where  $\omega$  and  $\eta$  are as defined in Section 4.2. We dispense with the normality assumptions on  $\omega$  and  $\eta$ . In the previous section, we established the risk optimality of the HB predictor  $e_t^{B*}(c_t^{(1)})$  within the class of all LUPs of  $\xi_t(C_t^{(1)}, C_t^{(2)})$  without any distributional assumptions on  $e^*$ . The optimality of  $e_t^{B*}(c_t^{(1)})$  within LUPs holds a fortiori under the quadratic loss

$$\begin{aligned} L_1(\xi_t, \delta) &= |\delta - \xi_t|_{\Omega}^2 \\ &= (\delta - \xi_t)^T \Omega (\delta - \xi_t) \\ &= \text{tr} [\Omega L_0(\xi_t, \delta)], \end{aligned}$$

where  $\Omega$  is a n.n.d. matrix. We shall refer to such a loss as generalized Euclidean error w.r.t.  $\Omega$ . Let  $R_L(\Phi, \xi_t; \delta) = E_{\Phi} \left[ L \left( \left| \delta(C_t^{(1)}) - \xi_t(C_t^{(1)}, C_t^{(2)}) \right|_{\Psi_t}^2 \right) \right]$  be the risk function of the predictor  $\delta$  for predicting  $\xi_t$  under a loss function which is a function of generalized Euclidean error w.r.t.  $\Omega$  for some function  $L$ .

We now assume that  $e^*$  has an elliptically symmetric distribution with pdf given by

$$h(e^* | \Delta, r) \propto |r^{-1} \Delta|^{-\frac{1}{2}} f(re^{*T} \Delta^{-1} e^*) \quad (4.5.24)$$

$$\Delta = \text{Block Diagonal} \left( B_t(I \otimes W)B_t^T, (I_m \otimes \Psi_t) + V_t \right)$$

and the known non negative function  $f$  is such that

$$\int \left[ \sum_{j=1}^t \sum_{k=1}^q |w_{jk}| + \sum_{i=1}^m \sum_{j=1}^t \sum_{k=1}^{N_{ij}} (|\varepsilon_{ijk}| + |e_{ijk}|) + 1 \right] f(re^{*T} \Delta^{-1} e^*) de^* < \infty \quad (4.5.25)$$

where

$$w_j = (w_{j1}, w_{j2}, \dots, w_{jk})^T$$

We denote this distribution by  $\mathfrak{F}_f(\mathbf{0}, r^{-1} \Delta)$  where  $\mathfrak{F}_f(\boldsymbol{\mu}, \boldsymbol{\Omega}^* \sigma^2)$  denotes the distribution whose pdf is given by

$$k(\mathbf{s} | \boldsymbol{\mu}, \boldsymbol{\Omega}^*, \sigma) \propto |\sigma^2 \boldsymbol{\Omega}^*|^{-\frac{1}{2}} f((\mathbf{s} - \boldsymbol{\mu})^T \boldsymbol{\Omega}^* (\mathbf{s} - \boldsymbol{\mu}) / \sigma^2) \quad (4.5.26)$$

where  $\mathbf{s}$  and  $\boldsymbol{\mu}$  are in  $R^p$ ,  $\boldsymbol{\Omega}^* (p \times p)$  is p.d. and  $\sigma > 0$ .

Note that the normality of  $\mathbf{e}$  with mean  $\mathbf{0}$  and variance covariance matrix  $r^{-1} \Delta$  is sufficient but not necessary for (4.5.24) and (4.5.25) to hold.

It follows from (4.5.25) that  $E(\mathbf{e}^*)$  exists and from (4.5.24)  $E(\mathbf{e}^*) = \mathbf{0}$

Note that  $\mathbf{e}^{**} = r^{-1} \Delta^{-\frac{1}{2}} \mathbf{e}^*$  has a spherically symmetric distribution  $\mathfrak{F}_f(\mathbf{0}, \mathbf{I}_{q^*})$  with characteristic function (c.f.)  $E[\exp(i\mathbf{u}^T \mathbf{e}^{**})] = c(\mathbf{u}^T \mathbf{u})$  for some function  $c$  (see Kelker, 1970) where  $i = \sqrt{-1}$ ,  $\mathbf{u} = (u_1, \dots, u_{q^*})^T$  and  $q^* = N_t + tq$ .

Hence  $\mathbf{e}^*$  has c.f. given by

$$E[\exp(i\mathbf{u}^T \mathbf{e}^*)] = c(r^{-1} \mathbf{u}^T \Delta \mathbf{u}) \quad (4.5.27)$$

Now write  $\mathbf{D}_{tj}^* = \mathbf{B}_t^{(j)} \mathbf{w} + \mathbf{e}^{(j)}$  ( $j = 1, 2$ )

Then  $\mathbf{D}_t^* = (\mathbf{D}_{t1}^{*T}, \mathbf{D}_{t2}^{*T})^T$  has a c.f. given by

$$E[\exp(i\mathbf{u}^{*T} \mathbf{D}_t^*)] = c(r^{-1} \mathbf{u}^{*T} \boldsymbol{\Sigma}_t \mathbf{u}^*) \quad (4.5.28)$$

where

$$\boldsymbol{\Sigma}_t = \mathbf{B}_t(\mathbf{I} \otimes \mathbf{W})\mathbf{B}_t^T + (\mathbf{I}_m \otimes \boldsymbol{\Psi}_t) + \mathbf{V}_t$$



Comparing (4.5.24) and (4.5.27) one can see from (4.5.28) that  $D_t^*$  has also an elliptically symmetric distributions  $\mathfrak{F}_f(0, r^{-1}\Sigma_t)$  with pdf given by

$$h(d^*|\Sigma_t, r) \propto |r^{-1}\Sigma_t|^{-\frac{1}{2}} f(r d_t^{*T} \Sigma_t^{-1} d_t^*) \quad (4.5.29)$$

We now introduce the notion of “stochastic” or “universal” domination as given in Hwang (1985).

**Definition 4.5.1.** An estimator  $\delta_1(C_t^{(1)})$  universally dominates  $\delta_2(C_t^{(1)})$  (under the generalized Euclidean error w.r.t  $\Omega$ ) if for every  $\Phi$ , and every nondecreasing loss function  $L$ ,  $R_L(\Phi, \xi_t; \delta_1) \leq R_L(\Phi, \xi_t; \delta_2)$  holds and for a particular loss, the risk functions are not identical.

In Hwang (1985), it has been shown that  $\delta_1$  universally dominates  $\delta_2$  under the generalized Euclidean error w.r.t  $\Omega$  if

$$|\delta_1(C_t^{(1)}) - \xi_t(C_t^{(1)}, C_t^{(2)})|_{\Omega}^2$$

is stochastically smaller than

$$|\delta_2(C_t^{(1)}) - \xi_t(C_t^{(1)}, C_t^{(2)})|_{\Omega}^2$$

The following theorem establishes the stochastic dominance of  $e_t^{B*}(C_t^{(1)})$  over every LUP of  $\xi_t(C_t^{(1)}, C_t^{(2)})$  for a general class of elliptically symmetric distributions of  $e^*$ .

**Theorem 4.5.1** Under the model given in 4.2.1 with  $\lambda$  known and assuming  $e^*$  has an elliptically symmetric pdf as in (4.5.24),  $e_t^{B*}(C_t^{(1)})$  universally dominates every LUP  $\delta(C_t^{(1)}) = S_t C_t^{(1)}$  of  $\xi_t(C_t^{(1)}, C_t^{(2)})$  for every p.d.  $\Omega$ .

To prove this theorem, we require the following lemma.

**Lemma 4.5.1** If  $D(N_t \times 1)$  has pdf  $h(d|I_{N_t}, r)$ , then for every  $L(u \times N_t)$ ,  $u \leq N_t$ ,  $L D \stackrel{d}{=} (L L^T)^{\frac{1}{2}} D_u = (I_u, 0) D$  where

$0(u \times (N_t - u))$  is a null matrix and  $\stackrel{d}{=}$  means equal in distribution.

Proof of Lemma 4.5.1 . The proof follows the arguments of Hwang (1985). From (4.5.27) it follows that  $D$  has c.f.  $E \left[ \exp \left( i s^T D \right) \right] = c \left( r^{-1} s^T s \right)$ . Hence

$$E \left[ \exp \left( i s_1^T L D \right) \right] = c \left( r^{-1} s_1^T L L^T s_1 \right) \quad (4.5.30)$$

where  $s_1$  is a  $u \times 1$  vector. Next using (4.5.30)

$$\begin{aligned} & E \left[ \exp \left( i s_1^T \left( L L^T \right)^{\frac{1}{2}} D_u \right) \right] \\ &= E \left[ \exp \left( i s_1^T \left( L L^T \right)^{\frac{1}{2}} \begin{pmatrix} I_u & 0 \end{pmatrix} D \right) \right] \\ &= c \left( r^{-1} s_1^T \left( L L^T \right)^{\frac{1}{2}} \begin{pmatrix} I_u & 0 \end{pmatrix} \begin{pmatrix} I_u & 0 \end{pmatrix}^T \left( L L^T \right)^{\frac{1}{2}} s_1 \right) \\ &= c \left( r^{-1} s_1^T \left( L L^T \right) s_1 \right), \end{aligned} \quad (4.5.31)$$

so that the lemma follows from (4.5.30) and (4.5.31).

It follows as a consequence of Lemma 4.5.1 that

$$\begin{aligned} D^T L^T L D &= (L D)^T (L D) \\ &\stackrel{d}{=} \left( \left( L L^T \right)^{\frac{1}{2}} D_u \right)^T \left( \left( L L^T \right)^{\frac{1}{2}} D_u \right) \\ &\stackrel{d}{=} D_u^T L L^T D_u \end{aligned} \quad (4.5.32)$$

We shall use (4.5.32) repeatedly for proving Theorem 4.5.1.

Proof of Theorem 4.5.1 Let  $S_t C_t^{(1)}$  be a LUP of  $\xi_t(C_t^{(1)}, C_t^{(2)})$ . Then, under the time series model (4.2.1) with  $\lambda$  known and from the fact that  $E(e^*) = 0$ , it is easy to see that

$$(S_t - P_1) A_t^{(1)} = P_2 A_t^{(2)} \quad (4.5.33)$$

Writing  $D = \Sigma_t^{-\frac{1}{2}} D^*$  and using (4.5.32) and (4.5.33) one gets

$$\begin{aligned}
 & [S_t C_t^{(1)} - \xi_t(C_t^{(1)}, C_t^{(2)})]^T \Omega [S_t C_t^{(1)} - \xi_t(C_t^{(1)}, C_t^{(2)})] \\
 &= D^{*T} [S_t - P_1 - P_2]^T \Omega [S_t - P_1 - P_2] D^* \\
 &= D^T \Sigma_t^{\frac{1}{2}} [S_t - P_1 - P_2]^T \Omega [S_t - P_1 - P_2] \Sigma_t^{\frac{1}{2}} D \\
 &\stackrel{d}{=} D_u^T \Omega^{\frac{1}{2}} [S_t - P_1 - P_2]^T \Sigma_t [S_t - P_1 - P_2] \Omega^{\frac{1}{2}} D_u.
 \end{aligned} \tag{4.5.34}$$

Similarly,

$$\begin{aligned}
 & [e_t^{B*}(C_t^{(1)}) - \xi_t(C_t^{(1)}, C_t^{(2)})]^T \Omega [e_t^{B*}(C_t^{(1)}) - \xi_t(C_t^{(1)}, C_t^{(2)})] \\
 &= D^{*T} [P_2 M_t - P_2]^T \Omega [P_2 M_t - P_2] D^* \\
 &= D^T \Sigma_t^{\frac{1}{2}} [P_2 M_t - P_2]^T \Omega [P_2 M_t - P_2] \Sigma_t^{\frac{1}{2}} D \\
 &\stackrel{d}{=} D_u^T \Omega^{\frac{1}{2}} [P_2 M_t - P_2] \Sigma_t [P_2 M_t - P_2]^T \Omega^{\frac{1}{2}} D_u.
 \end{aligned} \tag{4.5.35}$$

Write  $\Gamma = S_t - P_1 - P_2 M_t$ . Then

$$\text{rhs of (4.5.34)} = \text{rhs of (4.5.35)} + D_u^T \Omega^{\frac{1}{2}} \Gamma \Sigma_{t11} \Gamma^T \Omega^{\frac{1}{2}} D_u \tag{4.5.36}$$

since using the definition of  $M_t$ .

$$\begin{aligned}
 & (P_2 M_t - P_2) \Sigma_t \begin{pmatrix} \Gamma^T \\ 0 \end{pmatrix} \\
 &= P_2 (M_t - I) \begin{pmatrix} \Sigma_{t11} \\ \Sigma_{t21} \end{pmatrix} \Gamma^T \\
 &= P_2 (M_t \Sigma_{t11} - \Sigma_{t21}) \Gamma^T \\
 &= P_2 (A_t^{(2)} - \Sigma_{t21} \Sigma_{t11}^{-1} A_t^{(1)}) \left( A_t^{(1)T} \Sigma_{t11}^{-1} A_t^{(1)} \right)^{-1} A_t^{(1)T} \Gamma^T
 \end{aligned} \tag{4.5.37}$$

and using (4.5.33),

$$\Gamma A_t^{(1)} = (S_t - P_1 - P_2 M_t) A_t^{(1)}$$

$$\begin{aligned}
&= P_2 (A_t^{(2)} - M_t A_t^{(1)}) \\
&= 0
\end{aligned} \tag{4.5.38}$$

*Theorem (4.5.1) follows now from (4.5.34)–(4.5.36).*

*Also since  $\Sigma_{t11}$  is positive definite, it follows from (4.5.36) that*

$$\text{rhs of (59)} = \text{rhs of (60)}$$

*if and only if  $T = 0$ , that is  $S_t = P_1 + P_2 M_t$ .*

#### 4.6 Best Equivariant Prediction

In this section we concentrate on the equivariant prediction of  $\xi_t(C_t^{(1)}, C_t^{(2)})$  on the basis of  $C_t^{(1)}$  under suitable group of transformations. Assuming  $\lambda$  as known, we rewrite the model in (4.2.1) as

$$C_t \sim N(A_t \alpha, r^{-1} \Sigma_t) \tag{4.6.39}$$

where

$$\Sigma_t = B_t(I \otimes W)B_t^T + (I_m \otimes \Psi_t) + V_t \tag{4.6.40}$$

We redefine  $\phi = (\sigma \alpha^T)$

We are interested in equivariant estimation of  $P_1 A_t^{(1)} \alpha + P_2 A_t^{(2)} \alpha$  where the matrices  $A_t^{(1)}$ ,  $A_t^{(2)}$ ,  $P_1$  and  $P_2$  are as introduced earlier in Section 4.2. It suffices to estimate the vector  $\alpha$  where  $A_t \alpha$  can be viewed as a location vector.

Note that  $C_t + A_t \beta \sim N(A_t(\alpha + \beta), \sigma^2 \Sigma_t)$

We are interested in developing an estimator of  $\alpha$  which remains equivariant under a new origin of measurement. So a “natural” group of transformations will be

$$\mathcal{G}_1 = \{g\beta, \beta \in R^p : g\beta(C_t) = C_t + A_t \beta\} \tag{4.6.41}$$

Now assume that we partition  $C_t$  and  $A_t$  as

$$C_t = \begin{pmatrix} C_t^{(1)} \\ C_t^{(2)} \end{pmatrix}, A_t = \begin{pmatrix} A_t^{(1)} \\ A_t^{(2)} \end{pmatrix}$$

If  $\delta(C_t^{(1)})$  estimates  $(P_1 A_t^{(1)} + P_2 A_t^{(2)})\alpha$  then  $\delta(C_t^{(1)}) + P_1 A_t^{(1)}\beta + P_2 A_t^{(2)}\beta$  should estimate

$$(P_1 A_t^{(1)} + P_2 A_t^{(2)})(\alpha + \beta) = (P_1 \ P_2)A_t(\alpha + \beta)$$

Treating  $A_t(\alpha + \beta)$  as the new location parameter, one can expect that  $\delta(C_t^{(1)} + A_t^{(1)}\beta)$  (estimator based on the "observed part" of  $C_t$  and  $A_t$ ) will estimate

$$(P_1 A_t^{(1)} + P_2 A_t^{(2)})(\alpha + \beta)$$

So we should have

$$\delta(C_t^{(1)} + A_t^{(1)}\beta) = \delta(C_t^{(1)}) + P_1 A_t^{(1)}\beta + P_2 A_t^{(2)}\beta \quad (4.6.42)$$

for all  $C_t^{(1)}, \beta$ . Now if we are interested in estimating  $\xi_t(C_t^{(1)}, C_t^{(2)})$ , instead of

$$E_\phi(\xi_t(C_t^{(1)}, C_t^{(2)})) = P_1 A_t^{(1)}\alpha + P_2 A_t^{(2)}\alpha$$

We can still use  $\delta(C_t^{(1)})$  and again we will impose (4.6.42) on  $\delta$ .

Note that the induced group of transformations on the parameter space

$$\Phi = \{\phi : \phi = (\sigma, \alpha^T)^T, \alpha \in R^p, \sigma > 0\} \quad (4.6.43)$$

is given by

$$\bar{\mathcal{G}}_1 = \{\bar{g}\beta, \beta \in R^p : \bar{g}\beta(\phi) = (\sigma, \alpha^T + \beta^T)^T\} \quad (4.6.44)$$



A loss function  $L(\xi_t, \phi; \delta)$  is said to be invariant for predicting  $\xi_t(C_t^{(1)}, C_t^{(2)})$  by  $\delta(C_t^{(1)})$  if

$$\begin{aligned} & L(\xi_t(C_t^{(1)}, C_t^{(2)}) + (P_1 A_t^{(1)} + P_2 A_t^{(2)})\beta, \bar{g}_\beta(\phi); \\ & \delta(C_t^{(1)}) + (P_1 A_t^{(1)} + P_2 A_t^{(2)})\beta) \\ & = L(\xi_t(C_t^{(1)}, C_t^{(2)}), \phi; \delta(C_t^{(1)})) \end{aligned} \quad (4.6.45)$$

for all  $C_t^{(1)}, C_t^{(2)}, \beta$ , and  $\phi$ . An estimator  $\delta$  satisfying (4.6.42) is said to be equivariant. We will now be interested in the best equivariant prediction of

$$\xi_t(C_t^{(1)}, C_t^{(2)})$$

The following lemma provides a useful characterization of the class of equivariant predictors of  $\xi_t(C_t^{(1)}, C_t^{(2)})$ .

Lemma 4.6.1 . Let  $\delta_0(C_t^{(1)})$  be an equivariant predictor for  $\xi_t(C_t^{(1)}, C_t^{(2)})$ . Then a necessary and sufficient condition for a predictor  $\delta(C_t^{(1)})$  to be equivariant is that

$$\delta(C_t^{(1)}) = \delta_0(C_t^{(1)}) + h(C_t^{(1)}) \quad (4.6.46)$$

for all  $C_t^{(1)}$ , where

$$h(C_t^{(1)} + A_t^{(1)}\beta) = h(C_t^{(1)}) \quad (4.6.47)$$

for all  $C_t^{(1)}$  and all  $\beta$ .

A function  $h$  satisfying (4.6.47) is said to be invariant and hence must be a function of a maximal invariant function under the group of transformations

$$\mathcal{G}'_1 = \left\{ g'_\beta, \beta \in R^p : g'_\beta(C_t^{(1)}) = C_t^{(1)} + A_t^{(1)}\beta \right\} \quad (4.6.48)$$

induced by  $\mathcal{G}_1$  in the  $C_t^{(1)}$  - space. The next lemma finds a maximal invariant function under the group of transformations  $\mathcal{G}'_1$ .

**Lemma 4.6.2** *A maximal invariant function under the group of transformations  $\mathcal{G}'_1$  is  $K_t C_t^{(1)}$  where*

$$K_t = \Sigma_{t_{11}}^{-1} - \Sigma_{t_{11}}^{-1} A_t^{(1)} (A_t^{(1)T} \Sigma_{t_{11}}^{-1} A_t^{(1)})^{-1} A_t^{(1)T} \Sigma_{t_{11}}^{-1}$$

**Proof of Lemma 4.6.1** *First we show that  $K_t C_t^{(1)}$  is an invariant function. Note that*

$$K_t A_t^{(1)} = 0$$

*and hence*

$$K_t (C_t^{(1)} + A_t^{(1)} \beta) = K_t C_t^{(1)}$$

*for all  $C_t^{(1)}$  and  $\beta$ . Now we will prove the maximality.*

*For  $C_{t_1}^{(1)}, C_{t_2}^{(1)}$  such that*

$$K_t C_{t_1}^{(1)} = K_t C_{t_2}^{(1)}$$

*we have  $C_{t_1}^{(1)} - C_{t_2}^{(1)}$  is orthogonal to the row space of  $K_t$ .*

*Also  $\text{rank}(K_t) = n_t - p$  and  $\text{rank}(A_t^{(1)}) = p$ .*

*Again using  $K_t A_t^{(1)} = 0$ , it follows that the column space of  $A_t^{(1)}$  forms an orthocomplement of the row space of  $K_t$ .*

*Hence  $C_{t_1}^{(1)} - C_{t_2}^{(1)} = A_t^{(1)} \beta$  for some  $\beta$ , i.e.,*

$$C_{t_1}^{(1)} = C_{t_2}^{(1)} + A_t^{(1)} \beta$$

*proving thereby the maximality of  $K_t C_t^{(1)}$ .*

*Since  $e_t^{B*}(C_t^{(1)})$  is an equivariant predictor of  $\xi_t(C_t^{(1)}, C_t^{(2)})$ , using lemmas 4.6.1 and 4.6.2 we can characterize the class of all equivariant predictors of  $\xi_t(C_t^{(1)}, C_t^{(2)})$ . This is given in the next lemma.*

Lemma 4.6.3 Under the group of transformations  $\mathcal{G}_1$  given in (4.6.41), a predictor  $\delta(C_t^{(1)})$  of  $\xi_t(C_t^{(1)}, C_t^{(2)})$  is an equivariant predictor if and only if  $\delta$  has the representation

$$\delta(C_t^{(1)}) = e_t^{B*}(C_t^{(1)}) + l(K_t C_t^{(1)}) \quad (4.6.49)$$

for all  $C_t^{(1)}$ , where  $l$  is an  $n_t$ -variate vector valued function.

Define the loss functions

$$L_0(\xi_t, \delta) = (\delta - \xi_t)(\delta - \xi_t)^T, \quad L_1(\xi_t, \delta) = \text{tr}[\Omega L_0(\xi_t, \delta)], \quad (4.6.50)$$

where  $\Omega$  is a n.n.d. matrix.

Note that  $L_0$  and  $L_1$  are invariant loss functions.

Definition 4.6.1. An equivariant predictor  $\delta_0(C_t^{(1)})$  of  $\xi_t(C_t^{(1)}, C_t^{(2)})$  is said to be a best equivariant predictor under the loss  $L_0$  if for every other equivariant predictor  $\delta(C_t^{(1)})$  of  $\xi_t(C_t^{(1)}, C_t^{(2)})$

$$E_\phi [L_0(\xi_t, \delta) - L_0(\xi_t, \delta_0)]$$

is n.n.d.

Note that if  $\delta_0(C_t^{(1)})$  is the best equivariant predictor of  $\xi_t(C_t^{(1)}, C_t^{(2)})$  under the loss  $L_0$ , then it is so under the loss  $L_1$  for every n.n.d.  $\Omega$ . Conversely, if  $\delta_0(C_t^{(1)})$  is the best equivariant predictor of  $\xi_t(C_t^{(1)}, C_t^{(2)})$  under the loss  $L_1$  for every n.n.d.  $\Omega$ , then it is so under  $L_0$ . Next theorem establishes the optimality of  $e_t^{B*}(C_t^{(1)})$  within the class of all equivariant predictors  $\delta(C_t^{(1)})$  of  $\xi_t(C_t^{(1)}, C_t^{(2)})$  under the loss  $L_0$ , and a fortiori under the loss  $L_1$  for every n.n.d.  $\Omega$ .

Theorem 4.6.1 Under the normal time series linear model given in (4.2.1), the group of transformations  $\mathcal{G}_1$  given in (4.6.41) and the loss  $L_0$  given in (4.6.50), the best equivariant predictor of  $\xi_t(C_t^{(1)}, C_t^{(2)})$  is given by  $e_t^{B*}(C_t^{(1)})$ .

Proof of Theorem 4.6.1 Note that from (4.6.49) any equivariant predictor  $\delta(C_t^{(1)})$  of  $\xi_t(C_t^{(1)}, C_t^{(2)})$  is given by

$$\delta(C_t^{(1)}) = e_t^{B*}(C_t^{(1)}) + l(K_t C_t^{(1)})$$

Only those  $\delta$  are to be considered for which  $E_\phi[L_0(\xi_t, \delta)]$  is finite.

Note that  $E_\phi[L_0(\xi_t, e_t^{B*})]$  exists finitely and hence

$$E_\phi[l^T(K_t C_t^{(1)})l(K_t C_t^{(1)})] \leq 2tr \{E_\phi[L_0(\xi_t, e_t^{B*})] + E_\phi[L_0(\xi_t, \delta)]\}$$

Hence by Cauchy-Schwarz inequality  $E_\phi[l(K_t C_t^{(1)})l^T(K_t C_t^{(1)})]$  is finite.

Under the matrix loss  $L_0$ ,

$$\begin{aligned} E_\phi [L_0(\xi_t, \delta) - L_0(\xi_t, e_t^{B*})] \\ = E_\phi [l(K_t C_t^{(1)})l^T(K_t C_t^{(1)})] \\ + E_\phi [(e_t^{B*}(C_t^{(1)}) - \xi_t(C_t^{(1)}, C_t^{(2)}))l^T(K_t C_t^{(1)})] \end{aligned} \quad (4.6.51)$$

$$+ E_\phi [l(K_t C_t^{(1)})(e_t^{B*}(C_t^{(1)}) - \xi_t(C_t^{(1)}, C_t^{(2)}))^T] \quad (4.6.52)$$

Now we will show that the last two product terms are null matrices. Using the definitions of  $e_t^{B*}(C_t^{(1)})$  and  $M_t$ , we have

$$\begin{aligned} E_\phi [(e_t^{B*}(C_t^{(1)}) - \xi_t(C_t^{(1)}, C_t^{(2)}))l^T(K_t C_t^{(1)})] \\ = P_2 [M_t C_t^{(1)} - E_\phi(C_t^{(2)}|C_t^{(1)})] \\ = P_2 [M_t C_t^{(1)} - A_t^{(2)}\alpha - \Sigma_{t21}\Sigma_{t11}^{-1}(C_t^{(1)} - A_t^{(1)}\alpha)] \text{ a.e. Leb.} \\ = P_2 [A_t^{(2)} - \Sigma_{t21}\Sigma_{t11}^{-1}A_t^{(1)}] (A_t^{(1)T}\Sigma_{t11}^{-1}A_t^{(1)})^{-1} \\ \times (A_t^{(1)T}\Sigma_{t11}^{-1}C_t^{(1)} - A_t^{(1)T}\Sigma_{t11}^{-1}A_t^{(1)}\alpha) \end{aligned} \quad (4.6.53)$$

Now

$$\begin{aligned}
 \text{Cov}_\phi \left[ A_t^{(1)T} \Sigma_{t11}^{-1} C_t^{(1)}, K_t C_t^{(1)} \right] &= A_t^{(1)T} \Sigma_{t11}^{-1} \Sigma_{t11} K_t^T \\
 &= A_t^{(1)T} K_t \\
 &= 0
 \end{aligned}$$

Since  $C_t^{(1)}$  is multivariate normal and  $A_t^{(1)T} \Sigma_{t11}^{-1} C_t^{(1)}$  and  $K_t C_t^{(1)}$  are linear functions of  $C_t^{(1)}$ , so  $A_t^{(1)T} \Sigma_{t11}^{-1} C_t^{(1)}$  and  $K_t C_t^{(1)}$  are independently distributed. Now using (4.6.53) and independence of  $A_t^{(1)T} \Sigma_{t11}^{-1} C_t^{(1)}$  and  $K_t C_t^{(1)}$ , we have

$$\begin{aligned}
 &E_\phi \left[ \left( e_t^{B*} (C_t^{(1)}) - \xi_t (C_t^{(1)}, C_t^{(2)}) \right) l^T (K_t C_t^{(1)}) \right] \\
 &= E_\phi \left[ E_\phi \left\{ e_t^{B*} (C_t^{(1)}) - \xi_t (C_t^{(1)}, C_t^{(2)}) \mid C_t^{(1)} \right\} l^T (K_t C_t^{(1)}) \right] \\
 &= P_2 \left( A_t^{(2)} - \Sigma_{t21} \Sigma_{t11}^{-1} A_t^{(1)} \right) \left( A_t^{(1)T} \Sigma_{t11}^{-1} A_t^{(1)} \right)^{-1} \\
 &\quad \times E_\phi \left[ \left( A_t^{(1)T} \Sigma_{t11}^{-1} C_t^{(1)} - A_t^{(1)T} \Sigma_{t11}^{-1} A_t^{(1)} \alpha \right) l^T (K_t C_t^{(1)}) \right] \\
 &= P_2 \left( A_t^{(2)} - \Sigma_{t21} \Sigma_{t11}^{-1} A_t^{(1)} \right) \left( A_t^{(1)T} \Sigma_{t11}^{-1} A_t^{(1)} \right)^{-1} \\
 &\quad \times E_\phi \left( \left( A_t^{(1)T} \Sigma_{t11}^{-1} C_t^{(1)} - A_t^{(1)T} \Sigma_{t11}^{-1} A_t^{(1)} \alpha \right) E_\phi (l^T (K_t C_t^{(1)})) \right) \\
 &= 0
 \end{aligned} \tag{4.6.54}$$

Now from (4.6.52) and (4.6.54)

$$\begin{aligned}
 E_\phi \left[ L_0 (\xi_t, \delta) - L_0 (\xi_t, e_t^{B*}) \right] &= E_\phi \left[ l^T (K_t C_t^{(1)}) l^T (K_t C_t^{(1)}) \right] \\
 &\geq 0
 \end{aligned} \tag{4.6.55}$$

with equality if and only if

$$P_\phi \left[ l^T (K_t C_t^{(1)}) \right] = 0,$$



$$i.e., P_{\phi} [\delta (C_t^{(1)}) = e_t^{B*} (C_t^{(1)})] = 1$$

The proof of theorem is complete.

Next instead of  $\mathcal{G}_1$  we consider the group  $\mathcal{G}_2$  of transformations given by

$$\mathcal{G}_2 = \{g_{\beta,d}, \beta \in R^p, d > 0 : g_{\beta,d}(C_t) = d C_t + A_t \beta\} \quad (4.6.56)$$

A predictor  $\delta(C_t^{(1)})$  of  $\xi_t(C_t^{(1)}, C_t^{(2)})$  is said to equivariant under the group  $\mathcal{G}_2$  of transformations if

$$\delta(dC_t^{(1)} + A_t^{(1)}\beta) = d\delta(C_t^{(1)}) + (P_1 A_t^{(1)} + P_2 A_t^{(2)})\beta \quad (4.6.57)$$

for all  $C_t^{(1)}, \beta$  and  $d > 0$ . Note that  $e_t^{B*}(C_t^{(1)})$  is again an equivariant predictor of  $\xi_t(C_t^{(1)}, C_t^{(2)})$  under the above criterion. Also, note that if  $\delta$  is equivariant under  $\mathcal{G}_2$  taking  $d = 1$  in (4.6.57) it follows that it is equivariant under  $\mathcal{G}_1$ . So the class of equivariant predictors under  $\mathcal{G}_2$  is smaller than that under  $\mathcal{G}_1$ . But we will prove that  $e_t^{B*}(C_t^{(1)})$  is best inside this smaller class under a larger family of distributions of  $C_t$  which includes the normal distribution as a special case. We will apply the group of transformations  $\mathcal{G}_1$  on the family of elliptically symmetric distributions with pdf given by

$$f_{\phi}(c_t) \propto |\sigma^2 \Sigma_t|^{-\frac{1}{2}} f((c_t - A_t \alpha)^T \Sigma_t^{-1} (c_t - A_t \alpha) / \sigma^2) \quad (4.6.58)$$

where  $f$  is a known non-negative function satisfying

$$\int [1 + (c_t - A_t \alpha)^T \Sigma_t^{-1} (c_t - A_t \alpha)] \times f((c_t - A_t \alpha)^T \Sigma_t^{-1} (c_t - A_t \alpha) / \sigma^2) dc_t < \infty \quad (4.6.59)$$

Note that  $C_t \sim N(A_t \alpha, \sigma^2 \Sigma_t)$  is a special case satisfying (4.6.58) and (4.6.59). It is easy to see that the above group of transformations  $\mathcal{G}_2$  on  $N_t$ -dimensional Euclidean space for  $C_t$  induces a group of transformations

$$\bar{\mathcal{G}}_2 = \{\bar{g}_{\beta,d}, \beta \in R^p, d > 0 : \bar{g}_{\beta,d} = (d\sigma, d\alpha^T + \beta^T)^T\} \quad (4.6.60)$$

on the parameter space given in (4.6.43).

As before, a loss function  $L(\xi_t, \phi; \delta)$  for predicting  $\xi_t(C_t^{(1)}, C_t^{(2)})$  by  $\delta(C_t^{(1)})$  is invariant under the group of transformations  $\mathcal{G}_2$  if

$$\begin{aligned} & L(d\xi_t(C_t^{(1)}, C_t^{(2)}) + (P_1 A_t^{(1)} + P_2 A_t^{(2)})\beta, \bar{g}_{\beta,d}(\phi); \\ & \quad d\delta(C_t^{(1)}) + (P_1 A_t^{(1)} + P_2 A_t^{(2)})\beta) \\ & = L(\xi_t(C_t^{(1)}, C_t^{(2)}), \phi; \delta(C_t^{(1)})) \end{aligned} \quad (4.6.61)$$

for all  $C_t^{(1)}, C_t^{(2)}, \beta, d(> 0)$  and  $\phi$ .

We shall consider the special losses

$$L_2(\xi_t, \phi; \delta) = L_0(\xi_t, \delta) / \sigma^2 \quad (4.6.62)$$

and

$$L_3(\xi_t, \phi; \delta) = L_1(\xi_t, \delta) / \sigma^2 \quad (4.6.63)$$

Both these losses satisfy (4.6.61).

To find the best equivariant predictor of  $\xi_t(C_t^{(1)}, C_t^{(2)})$  in this set up, we will present two lemmas which provide a useful characterization of the class of equivariant predictors.

Lemma 4.6.4 *Let  $\delta_0(C_t^{(1)})$  be an equivariant predictor for  $\xi_t(C_t^{(1)}, C_t^{(2)})$ . Then a necessary and sufficient condition for a predictor  $\delta(C_t^{(1)})$  of  $\xi_t(C_t^{(1)}, C_t^{(2)})$  to be equivariant is that*

$$\delta(C_t^{(1)}) = \delta_0(C_t^{(1)}) + h(C_t^{(1)}) \quad (4.6.64)$$

for all  $C_t^{(1)}$ , where

$$h(dC_t^{(1)} + A_t^{(1)}\beta) = dh(C_t^{(1)}) \quad (4.6.65)$$

for all  $C_t^{(1)}, \beta$  and  $d > 0$ .

We will follow the arguments of Lemma 2 and 3 of Datta and Ghosh (1988) to find a representation of the function  $h$  in (4.6.65). Define

$$p(K_t C_t^{(1)}) = K_t C_t^{(1)} / (C_t^{(1)T} K_t C_t^{(1)})^{\frac{1}{2}} I_{[C_t^{(1)T} K_t C_t^{(1)} > 0]} \quad (4.6.66)$$

Note that since  $K_t \Sigma_{t11} K_t = K_t$ , so

$$C_t^{(1)T} K_t C_t^{(1)} = (K_t C_t^{(1)})^T \Sigma_{t11} (K_t C_t^{(1)})$$

and  $p$  is indeed a function of  $K_t C_t^{(1)}$ . It can be shown that  $p(K_t C_t^{(1)})$  is a maximal invariant under the group of transformations

$$\mathcal{G}'_2 = \left\{ g'_{\beta,d}, \beta \in R^p, d > 0 : g'_{\beta,d}(C_t^{(1)}) = (dC_t^{(1)} + A_t^{(1)})\beta \right\} \quad (4.6.67)$$

induced by  $\mathcal{G}_2$  in the  $C_t^{(1)}$ -space.

The following lemma characterizes the class of functions  $h(C_t^{(1)})$  satisfying (4.6.65). We will use this lemma to characterize the class of equivariant predictors.

Lemma 4.6.5 A function  $h(C_t^{(1)})(u \times 1)$  satisfying (4.6.65) if and only if  $h$  has the representation

$$h(C_t^{(1)}) = (C_t^{(1)T} K_t C_t^{(1)})^{\frac{1}{2}} s(p(K_t C_t^{(1)})) \quad (4.6.68)$$

where  $s(u \times 1)$  is an arbitrary function of  $p(K_t C_t^{(1)})$ .

Proof of Lemma 4.6.2 Assume  $h$  has the representation given in (4.6.68). Now, since  $(dC_t^{(1)} + A_t^{(1)}\beta)^T K_t (dC_t^{(1)} + A_t^{(1)}\beta) = d^2 C_t^{(1)T} K_t C_t^{(1)}$  and since  $p(K_t C_t^{(1)})$  is a maximal invariant under  $\mathcal{G}'_2$ , so  $p(K_t (dC_t^{(1)} + A_t^{(1)}\beta)) = p(K_t C_t^{(1)})$  and consequently

$$\begin{aligned} h(dC_t^{(1)} + A_t^{(1)}\beta) &= (d^2 C_t^{(1)T} K_t C_t^{(1)})^{\frac{1}{2}} s(p(K_t C_t^{(1)})) \\ &= dh(C_t^{(1)}) \end{aligned}$$

Hence (4.6.65) is satisfied. only if. Since  $\mathbf{h}$  satisfies (4.6.65) for all  $\mathbf{c}_t^{(1)}, \beta$  and  $d > 0$ , taking  $d = 1$ , we see that  $\mathbf{h}$  must satisfy  $\mathbf{h}(\mathbf{C}_t^{(1)} + \mathbf{A}_t^{(1)}\beta) = \mathbf{h}(\mathbf{C}_t^{(1)})$  for all  $\mathbf{C}_t^{(1)}$  and  $\beta$ . This implies that  $\mathbf{h}$  must be invariant under  $\mathcal{G}'_1$ , and hence must be a function of  $\mathbf{K}_t\mathbf{C}_t^{(1)}$ . So  $\mathbf{h}(\mathbf{C}_t^{(1)}) = \mathbf{s}(\mathbf{K}_t\mathbf{C}_t^{(1)})$  where  $\mathbf{s}(u \times 1)$  is an arbitrary function satisfying

$$\begin{aligned} \mathbf{s}(\mathbf{K}_t(d\mathbf{C}_t^{(1)} + \mathbf{A}_t^{(1)}\beta)) &= d\mathbf{s}(\mathbf{K}_t\mathbf{C}_t^{(1)}) \\ \text{i.e. } \mathbf{s}(d\mathbf{K}_t\mathbf{C}_t^{(1)}) &= d\mathbf{s}(\mathbf{K}_t\mathbf{C}_t^{(1)}) \end{aligned} \quad (4.6.69)$$

for all  $d > 0$ . Now taking  $d = (\mathbf{C}_t^{(1)T} \mathbf{K}_t \mathbf{C}_t^{(1)})^{-\frac{1}{2}}$  for all  $\mathbf{C}_t^{(1)T} \mathbf{K}_t \mathbf{C}_t^{(1)} > 0$  we have from (4.6.69)

$$\begin{aligned} \mathbf{h}(\mathbf{C}_t^{(1)}) &= \mathbf{s}(\mathbf{K}_t\mathbf{C}_t^{(1)}) \\ &= \mathbf{s} \left( \left( \mathbf{C}_t^{(1)T} \mathbf{K}_t \mathbf{C}_t^{(1)} \right)^{-\frac{1}{2}} \mathbf{K}_t \mathbf{C}_t^{(1)} \right) \left( \mathbf{C}_t^{(1)T} \mathbf{K}_t \mathbf{C}_t^{(1)} \right)^{\frac{1}{2}} \\ &= \mathbf{s} \left( \mathbf{p} \left( \mathbf{K}_t \mathbf{C}_t^{(1)} \right) \right) \left( \mathbf{C}_t^{(1)T} \mathbf{K}_t \mathbf{C}_t^{(1)} \right)^{\frac{1}{2}} \end{aligned}$$

Now for  $\mathbf{C}_t^{(1)T} \mathbf{K}_t \mathbf{C}_t^{(1)} = 0$ , if we take  $\mathbf{h} = \mathbf{0}$ , then (4.6.65) is satisfied and we can represent  $\mathbf{h}$  by (4.6.68).

Since  $\mathbf{e}_t^{B*}(\mathbf{C}_t^{(1)})$  is an equivariant predictor of  $\xi_t(\mathbf{C}_t^{(1)}, \mathbf{C}_t^{(2)})$ , it follows from lemma and lemma that  $\delta(\mathbf{C}_t^{(1)})$  is an equivariant predictor of  $\xi_t(\mathbf{C}_t^{(1)}, \mathbf{C}_t^{(2)})$  under the group  $\mathcal{G}_2$  of transformations if and only if

$$\delta(\mathbf{C}_t^{(1)}) = \mathbf{e}_t^{B*}(\mathbf{C}_t^{(1)}) + (\mathbf{C}_t^{(1)T} \mathbf{K}_t \mathbf{C}_t^{(1)})^{\frac{1}{2}} \mathbf{s}(\mathbf{p}(\mathbf{K}_t \mathbf{C}_t^{(1)})) \quad (4.6.70)$$

for all  $\mathbf{C}_t^{(1)}$ .

Definition An equivariant predictor  $\delta_0(\mathbf{C}_t^{(1)})$  of  $\xi_t(\mathbf{C}_t^{(1)}, \mathbf{C}_t^{(2)})$  is said to be a best equivariant predictor under the loss  $L_2$  if for every other equivariant predictor  $\delta(\mathbf{C}_t^{(1)})$  of  $\xi_t(\mathbf{C}_t^{(1)}, \mathbf{C}_t^{(2)})$

$$E_{\phi} [L_2(\xi_t, \phi; \delta) - L_2(\xi_t, \phi; \delta_0)]$$

is non negative definite.

Remark. Note that if  $\delta_0(C_t^{(1)})$  is the best equivariant predictor of  $\xi_t(C_t^{(1)}, C_t^{(2)})$  under the loss  $L_2$ , then it is so under the loss  $L_3$  for every n.n.d.  $\Omega$  and vice versa.

We now establish that  $e_t^{B*}(C_t^{(1)})$  is best equivariant predictor of  $\xi_t(C_t^{(1)}, C_t^{(2)})$  under the loss  $L_2$ .

Theorem 4.6.2 Under the time domain HB model given in and the family of elliptically symmetric distributions as given in (4.6.58) and (4.6.59), the group of transformations  $\mathcal{G}_2$  given in (4.6.56), and the loss function  $L_2$  given in (4.6.62), the best equivariant predictor of  $\xi_t(C_t^{(1)}, C_t^{(2)})$  is given by  $e_t^{B*}(C_t^{(1)})$ .

Proof of Theorem 4.6.2 Note that from (4.6.70) any equivariant predictor  $\delta(C_t^{(1)})$  of  $\xi_t(C_t^{(1)}, C_t^{(2)})$  is given by

$$\delta(C_t^{(1)}) = e_t^{B*}(C_t^{(1)}) + (C_t^{(1)T} K_t C_t^{(1)})^{\frac{1}{2}} s(p(K_t C_t^{(1)})).$$

Only those  $\delta$  are considered for which  $E_{\phi} (L_2(\xi_t, \phi; \delta))$  is finite and in that case for the corresponding  $s$ , we have

$$E_{\phi} \left[ C_t^{(1)T} K_t C_t^{(1)} s(p(K_t C_t^{(1)})) s^T(p(K_t C_t^{(1)})) \right]$$

is finite. Under the matrix loss  $L_2$ ,

$$\begin{aligned} & \sigma^2 E_{\phi} [L_2(\xi_t, \phi; \delta) - (\xi_t, \phi; e_t^{B*})] \\ &= E_{\phi} \left[ C_t^{(1)T} K_t C_t^{(1)} s(p(K_t C_t^{(1)})) s^T(p(K_t C_t^{(1)})) \right] \\ &+ E_{\phi} \left[ (e_t^{B*}(C_t^{(1)}) - \xi_t(C_t^{(1)}, C_t^{(2)})) (C_t^{(1)T} K_t C_t^{(1)})^{\frac{1}{2}} s^T(p(K_t C_t^{(1)})) \right] \\ &+ E_{\phi} \left[ (C_t^{(1)T} K_t C_t^{(1)})^{\frac{1}{2}} s(p(K_t C_t^{(1)})) (e_t^{B*}(C_t^{(1)}) - \xi_t(C_t^{(1)}, C_t^{(2)}))^T \right] \quad (4.6.71) \end{aligned}$$

It is enough to prove that the last two product terms in (4.6.71) are null matrices.



$$E_{\phi} [e_t^{B*}(C_t^{(1)}) - \xi_t(C_t^{(1)}, C_t^{(2)})|C_t^{(1)}] = P_2 [M_t C_t^{(1)} - E_{\phi}(C_t^{(2)}|C_t^{(1)})] \quad (4.6.72)$$

Now from the property of elliptically symmetric distributions, it follows that

$$E_{\phi}[C_t^{(2)}|C_t^{(1)}] = A_t^{(2)}\alpha + \Sigma_{t21}\Sigma_{t11}^{-1}(C_t^{(1)} - A_t^{(1)}\alpha) \text{ a.e. Lebesgue}$$

and using this, we have from (4.6.72) that

$$\begin{aligned} E_{\phi} [e_t^{B*}(C_t^{(1)}) - \xi_t(C_t^{(1)}, C_t^{(2)})|C_t^{(1)}] \\ = P_2 (A_t^{(2)} - \Sigma_{t21}\Sigma_{t11}^{-1}A_t^{(1)}) (A_t^{(1)T}\Sigma_{t11}^{-1}A_t^{(1)})^{-1} \\ \times (A_t^{(1)T}\Sigma_{t11}^{-1}C_t^{(1)} - A_t^{(1)T}\Sigma_{t11}^{-1}A_t^{(1)}\alpha) \end{aligned} \quad (4.6.73)$$

Then, the second term of the rhs of (4.6.71)

$$\begin{aligned} = P_2 (A_t^{(2)} - \Sigma_{t21}\Sigma_{t11}^{-1}A_t^{(1)}) (A_t^{(1)T}\Sigma_{t11}^{-1}A_t^{(1)})^{-1} \\ \times E_{\phi} \left[ (C_t^T K_t C_t)^{\frac{1}{2}} (A_t^{(1)T}\Sigma_{t11}^{-1}C_t^{(1)} - A_t^{(1)T}\Sigma_{t11}^{-1}A_t^{(1)}\alpha) \right. \\ \left. \times s^T(p(K_t C_t^{(1)})) \right] \end{aligned} \quad (4.6.74)$$

Now we will show that the expectation of the rhs of (4.6.73) is a null matrix. Note that since  $E_{\Phi}(C_t^{(1)T}K_t C_t^{(1)}) < \infty$  and  $E_{\Phi}(C_t^{(1)}) = A_t^{(1)}\alpha$ , by Cauchy-Schwarz inequality,  $(C_t^{(1)T}K_t C_t^{(1)})^{\frac{1}{2}}A_t^{(1)T}\Sigma_{t11}^{-1}A_t^{(1)}$  has finite expectation. Now using the independence of  $((C_t^{(1)T}K_t C_t^{(1)})^{\frac{1}{2}}, A_t^{(1)T}\Sigma_{t11}^{-1}C_t^{(1)})$  and  $p(K_t C_t^{(1)})$ , we have

$$\begin{aligned} E_{\phi} \left[ (C_t^T K_t C_t)^{\frac{1}{2}} (A_t^{(1)T}\Sigma_{t11}^{-1}C_t^{(1)} - A_t^{(1)T}\Sigma_{t11}^{-1}A_t^{(1)}\alpha) \right. \\ \left. \times s^T(p(K_t C_t^{(1)})) \right] \\ E_{\phi} \left[ (C_t^T K_t C_t)^{\frac{1}{2}} A_t^{(1)T}\Sigma_{t11}^{-1}(C_t^{(1)} - A_t^{(1)}\alpha) \right] \\ \times E_{\phi} [s^T(p(K_t C_t^{(1)}))] \end{aligned} \quad (4.6.75)$$

Next using the elliptical symmetry of  $C_t^{(1)}$ , it follows that

$$C_t^{(1)} - A_t^{(1)}\alpha \stackrel{d}{=} -(C_t^{(1)} - A_t^{(1)}\alpha)$$

Now since  $K_t A_t^{(1)} = 0$  and  $K_t$  is n.n.d., one has

$$\begin{aligned} & \left[ (C_t^{(1)} - A_t^{(1)}\alpha)^T K_t (C_t^{(1)} - A_t^{(1)}\alpha) \right]^{\frac{1}{2}} \left[ A_t^{aT} \Sigma_{t11}^{-1} (C_t^{(1)} - A_t^{(1)}\alpha) \right] \\ &= \left[ \left\{ -(C_t^{(1)} - A_t^{(1)}\alpha) \right\}^T K_t \left\{ -(C_t^{(1)} - A_t^{(1)}\alpha) \right\} \right]^{\frac{1}{2}} \\ & \times A_t^{aT} \Sigma_{t11}^{-1} \left\{ -(C_t^{(1)} - A_t^{(1)}\alpha) \right\} \\ & i.e. (C_t^T K_t C_t)^{\frac{1}{2}} A_t^{aT} (C_t^{(1)} - A_t^{(1)}\alpha) \\ &= -(C_t^{(1)T} K_t C_t^{(1)})^{\frac{1}{2}} A_t^{aT} \Sigma_{t11}^{-1} (C_t^{(1)} - A_t^{(1)}\alpha) \end{aligned} \quad (4.6.76)$$

Now taking expectations on both sides which exist finitely due to the previous observation, one gets

$$E_\phi \left[ (C_t^T K_t C_t)^{\frac{1}{2}} A_t^{aT} \Sigma_{t11}^{-1} (C_t^{(1)} - A_t^{(1)}\alpha) \right] = 0 \quad (4.6.77)$$

Hence from (4.6.71), (4.6.74), (4.6.75), and (4.6.77) we have

$$\begin{aligned} & \sigma^2 E_\phi \left[ L_2(\xi_t, \phi; \delta) - (\xi_t, \phi; e_t^{B*}) \right] \\ &= E_\phi \left[ (C_t^{(1)T} K_t C_t^{(1)}) s(p(K_t C_t^{(1)})) s^T(p(K_t C_t^{(1)})) \right] \end{aligned}$$

which is n.n.d. for all  $\phi$  and the two risk matrices are equal if and only if

$$\begin{aligned} & P_\phi \left[ s(p(K_t C_t^{(1)})) = 0 \right] = 1 \text{ for all } \phi \\ & i.e., P_\phi \left[ \delta(C_t^{(1)}) = e_t^{B*}(C_t^{(1)}) \right] = 1 \text{ for all } \phi \end{aligned}$$

Remark. Although  $(C_t^{(1)T} K_t C_t^{(1)})^{\frac{1}{2}}, A_t^{(1)T} \Sigma_{t11}^{-1} C_t^{(1)}$  is sufficient and  $p(K_t C_t^{(1)})$  is ancillary, Basu's theorem can not be applied since the sufficient statistics is not complete.

## CHAPTER 5

### SUMMARY AND FUTURE RESEARCH

#### 5.1 Summary

In this dissertation, an estimation and prediction methodology is developed for simultaneous estimation of several strata means at different time points, based on data from repeated surveys. The methodology adopts a HB model which incorporates time dependent random components. The resulting HB estimators for a small area “borrow strength” over time in addition to borrowing strength from other local areas as has been the case in most of the existing methods of small area estimation. The Bayesian procedure provides current as well as retrospective estimates simultaneously for all the small areas. Application of this HB procedure to the median income estimation problem, in Chapters 2 and 3 has demonstrated the usefulness of this procedure in providing substantially improved estimators for small areas.

Implementation of the HB methodology, in finite as well as in infinite population sampling situations have been illustrated by adopting a Monto Carlo integration technique known as the Gibbs sampler. Using this procedure, in both the univariate and the multivariate cases, the posterior density as well as conditional mean and variance can be obtained with considerable ease.

The HB predictors of a vector of linear functions of the finite population observations, in the special case of known ratios of variance components, are shown to be best unbiased linear predictors without any distributional assumptions. It is also

shown that these predictors stochastically dominate all the linear unbiased predictors for elliptically symmetric distributions and are best equivariant predictors under suitable group of transformations.

## 5.2 Future Research

The HB analysis considered in this dissertation assumed the variance-covariance matrix for the basic observations as known. However, often in practice, only the sample variance-covariance matrix is available. We can extend our HB model by setting either proper or improper priors on the unknown variance-covariance matrices. We also assumed that the process through which the random components evolve through time is known. The case when this process is unknown needs further investigation. Also, the application of general technique of “borrowing strength” over time can be explored in other application areas e.g. in multicenter clinical trials in biomedical research or in on-line quality assessment of different production lines in a manufacturing industry.

It is important to investigate the robustness of the HB procedure considered in this dissertation, to handle the case of heavy tailed priors. It is also, worthwhile to extend our Bayesian model to allow for the presence of outliers in the data by assuming scale-contaminated normal distributions.

APPENDIX  
PROOF OF THEOREM 4.2.1.

The joint pdf of  $C_t$ ,  $\alpha$ ,  $R$  and  $\Lambda$  is given by

$$f(c_t, \alpha, r, \lambda) \propto r^{\frac{1}{2}(N_t + \sum_{l=1}^s h_l) - 1} \left[ \prod_{l=1}^s \lambda_l^{\frac{1}{2}h_l - 1} \right] |\Sigma_t|^{-\frac{1}{2}} \times \exp \left[ -\frac{1}{2} \left\{ (c_t - A_t \alpha)^T \Sigma_t^{-1} (c_t - A_t \alpha) + a_0 + \sum_{l=1}^s a_l \lambda_l \right\} \right] \quad (1)$$

$$(c_t - A_t \alpha)^T \Sigma_t^{-1} (c_t - A_t \alpha) = \left[ \alpha - (A_t^T \Sigma_t^{-1} A_t)^{-1} A_t^T \Sigma_t^{-1} c_t \right]^T \times (A_t^T \Sigma_t^{-1} A_t) \left[ \alpha - (A_t^T \Sigma_t^{-1} A_t)^{-1} A_t^T \Sigma_t^{-1} c_t \right] + c_t^T Q_t c_t \quad (2)$$

where

$$Q_t = \Sigma_t^{-1} - \Sigma_t^{-1} A_t (A_t^T \Sigma_t^{-1} A_t)^{-1} A_t^T \Sigma_t^{-1}$$

Integrating (1) w.r.t  $\alpha$ , the joint pdf of  $C_t$ ,  $R$  and  $\Lambda$  is given by

$$f(c_t, r, \lambda) \propto |\Sigma_t|^{-\frac{1}{2}} |A_t^T \Sigma_t^{-1} A_t|^{-\frac{1}{2}} r^{\frac{1}{2}(N_t + \sum_{l=1}^s h_l - p)} \times \exp \left[ -\frac{1}{2} \left( a_0 + \sum_{l=1}^s a_l \lambda_l + c_t^T Q_t c_t \right) \right] \prod_{l=1}^s \lambda_l^{\frac{1}{2}h_l - 1} \quad (3)$$

Integrating (3) w.r.t  $R$ , the pdf of  $C_t$ , and  $\Lambda$  is given by

$$f(c_t, \lambda) \propto |\Sigma_t|^{-\frac{1}{2}} |A_t^T \Sigma_t^{-1} A_t|^{-\frac{1}{2}} \left( a_0 + \sum_{l=1}^s a_l \lambda_l + c_t^T Q_t c_t \right)^{\frac{1}{2}(N_t + \sum_{l=0}^s h_l - p)}$$



$$\times \prod_{l=1}^s \lambda_l^{\frac{1}{2}h_l-1} \quad (4)$$

Now, using a standard formula for partitioned matrices (e.g., Searle (1971), page 46), we have

$$\begin{aligned} \mathbf{c}_t^T \Sigma_t^{-1} \mathbf{c}_t &= \mathbf{c}_t^{(1)T} \Sigma_{t11}^{-1} \mathbf{c}_t^{(1)} + (\mathbf{c}_t^{(2)} - \Sigma_{t21} \Sigma_{t11}^{-1} \mathbf{c}_t^{(1)})^T \Sigma_{t22.1}^{-1} \\ &\quad \times (\mathbf{c}_t^{(2)} - \Sigma_{t21} \Sigma_{t11}^{-1} \mathbf{c}_t^{(1)}) \end{aligned} \quad (5)$$

Similarly

$$\begin{aligned} \mathbf{c}_t^T \Sigma_t^{-1} \mathbf{A}_t &= \mathbf{c}_t^{(1)T} \Sigma_{t11}^{-1} \mathbf{A}_t^{(1)} + (\mathbf{c}_t^{(2)} - \Sigma_{t21} \Sigma_{t11}^{-1} \mathbf{A}_t^{(1)})^T \Sigma_{t22.1}^{-1} \\ &\quad \times (\mathbf{c}_t^{(2)} - \Sigma_{t21} \Sigma_{t11}^{-1} \mathbf{A}_t^{(1)}) \\ &= \mathbf{s}_1^T + \mathbf{s}_2^T, \end{aligned} \quad (6)$$

$$\begin{aligned} \mathbf{A}_t^T \Sigma_t^{-1} \mathbf{A}_t &= \mathbf{A}_t^{(1)T} \Sigma_{t11}^{-1} \mathbf{A}_t^{(1)} + (\mathbf{A}_t^{(2)} - \Sigma_{t21} \Sigma_{t11}^{-1} \mathbf{A}_t^{(1)})^T \Sigma_{t22.1}^{-1} \\ &\quad \times (\mathbf{A}_t^{(2)} - \Sigma_{t21} \Sigma_{t11}^{-1} \mathbf{A}_t^{(1)}) \end{aligned} \quad (7)$$

From (7)

$$\begin{aligned} (\mathbf{A}_t^T \Sigma_t^{-1} \mathbf{A}_t)^{-1} &= \\ &= (\mathbf{A}_t^{(1)T} \Sigma_{t11}^{-1} \mathbf{A}_t^{(1)})^{-1} - (\mathbf{A}_t^{(1)T} \Sigma_{t11}^{-1} \mathbf{A}_t^{(1)})^{-1} (\mathbf{A}_t^{(2)} - \Sigma_{t21} \Sigma_{t11}^{-1} \mathbf{A}_t^{(1)})^T \\ &\quad \times \left( \Sigma_{t22.1} + (\mathbf{A}_t^{(2)} - \Sigma_{t21} \Sigma_{t11}^{-1} \mathbf{A}_t^{(1)}) (\mathbf{A}_t^{(1)T} \Sigma_{t11}^{-1} \mathbf{A}_t^{(1)})^{-1} \right. \\ &\quad \left. (\mathbf{A}_t^{(2)} - \Sigma_{t21} \Sigma_{t11}^{-1} \mathbf{A}_t^{(1)})^T \right)^{-1} \\ &\quad \times (\mathbf{A}_t^{(2)} - \Sigma_{t21} \Sigma_{t11}^{-1} \mathbf{A}_t^{(1)}) (\mathbf{A}_t^{(1)T} \Sigma_{t11}^{-1} \mathbf{A}_t^{(1)})^{-1} \\ &= (\mathbf{A}_t^{(1)T} \Sigma_{t11}^{-1} \mathbf{A}_t^{(1)})^{-1} - (\mathbf{A}_t^{(1)T} \Sigma_{t11}^{-1} \mathbf{A}_t^{(1)})^{-1} (\mathbf{A}_t^{(2)} - \Sigma_{t21} \Sigma_{t11}^{-1} \mathbf{A}_t^{(1)})^T \\ &\quad \times \mathbf{G}_t^{-1} (\mathbf{A}_t^{(2)} - \Sigma_{t21} \Sigma_{t11}^{-1} \mathbf{A}_t^{(1)}) (\mathbf{A}_t^{(1)T} \Sigma_{t11}^{-1} \mathbf{A}_t^{(1)})^{-1} \\ &= \mathbf{M}_1 + \mathbf{M}_2 \text{ (say)} \end{aligned} \quad (8)$$

Thus

$$\begin{aligned}
 \mathbf{c}_t^T \mathbf{Q}_t \mathbf{c}_t &= \mathbf{c}_t^T \boldsymbol{\Sigma}_t^{-1} \mathbf{A}_t (\mathbf{A}_t^T \boldsymbol{\Sigma}_t^{-1} \mathbf{A}_t)^{-1} \mathbf{A}_t^T \boldsymbol{\Sigma}_t^{-1} \mathbf{c}_t \\
 &= (\mathbf{s}_1^T + \mathbf{s}_2^T)(\mathbf{M}_1 + \mathbf{M}_2)(\mathbf{s}_1 + \mathbf{s}_2) \\
 &= \mathbf{s}_1^T \mathbf{M}_1 \mathbf{s}_1 - \mathbf{s}_1^T \mathbf{M}_2 \mathbf{s}_1 + \mathbf{s}_2^T \mathbf{M}_1 \mathbf{s}_2 \\
 &\quad - \mathbf{s}_2^T \mathbf{M}_2 \mathbf{s}_2 + 2\mathbf{s}_1^T (\mathbf{M}_1 - \mathbf{M}_2) \mathbf{s}_2
 \end{aligned} \tag{9}$$

$$\mathbf{s}_1^T \mathbf{M}_1 \mathbf{s}_1 = \mathbf{c}_t^{(1)T} (\boldsymbol{\Sigma}_{t11}^{-1} - \mathbf{K}_t) \mathbf{c}_t^{(1)} \tag{10}$$

$$\begin{aligned}
 \mathbf{s}_1^T \mathbf{M}_2 \mathbf{s}_2 &= (\mathbf{M}_t \mathbf{c}_t^{(1)} - \boldsymbol{\Sigma}_{t21} \boldsymbol{\Sigma}_{t11}^{-1} \mathbf{c}_t^{(1)})^T \mathbf{G}_t^{-1} (\mathbf{M}_t \mathbf{c}_t^{(1)} - \boldsymbol{\Sigma}_{t21} \boldsymbol{\Sigma}_{t11}^{-1} \mathbf{c}_t^{(1)}) \\
 \mathbf{s}_2^T \mathbf{M}_1 \mathbf{s}_2 &= (\mathbf{c}_t^{(2)} - \boldsymbol{\Sigma}_{t21} \boldsymbol{\Sigma}_{t11}^{-1} \mathbf{A}_t^{(1)})^T (\boldsymbol{\Sigma}_{t22.1}^{-1} \mathbf{G}_t \boldsymbol{\Sigma}_{t22.1}^{-1} - \boldsymbol{\Sigma}_{t22.1}^{-1}) \\
 &\quad \times (\mathbf{c}_t^{(2)} - \boldsymbol{\Sigma}_{t21} \boldsymbol{\Sigma}_{t11}^{-1} \mathbf{A}_t^{(1)})
 \end{aligned} \tag{11}$$

$$\begin{aligned}
 \mathbf{s}_2^T \mathbf{M}_2 \mathbf{s}_2 &= (\mathbf{c}_t^{(2)} - \boldsymbol{\Sigma}_{t21} \boldsymbol{\Sigma}_{t11}^{-1} \mathbf{c}_t^{(1)})^T (\boldsymbol{\Sigma}_{t22.1}^{-1} \mathbf{G}_t \boldsymbol{\Sigma}_{t22.1}^{-1} - 2\boldsymbol{\Sigma}_{t22.1}^{-1} \mathbf{G}_t^{-1}) \\
 &\quad \times (\mathbf{c}_t^{(2)} - \boldsymbol{\Sigma}_{t21} \boldsymbol{\Sigma}_{t11}^{-1} \mathbf{c}_t^{(1)})
 \end{aligned} \tag{12}$$

$$\mathbf{s}_1^T \mathbf{M}_1 \mathbf{s}_2 = (\mathbf{M}_t \mathbf{c}_t^{(1)} - \boldsymbol{\Sigma}_{t21} \boldsymbol{\Sigma}_{t11}^{-1} \mathbf{c}_t^{(1)})^T \boldsymbol{\Sigma}_{t22.1}^{-1} (\mathbf{c}_t^{(2)} - \boldsymbol{\Sigma}_{t21} \boldsymbol{\Sigma}_{t11}^{-1} \mathbf{c}_t^{(1)}) \tag{13}$$

$$\mathbf{c}_t^T \mathbf{Q}_t \mathbf{c}_t = \mathbf{c}_t^{(1)T} \mathbf{K}_t \mathbf{c}_t^{(1)} + (\mathbf{c}_t^{(2)} - \mathbf{M}_t \mathbf{c}_t^{(1)})^T \mathbf{G}_t^{-1} (\mathbf{c}_t^{(2)} - \mathbf{M}_t \mathbf{c}_t^{(1)}) \tag{14}$$

Now starting with the joint pdf of  $\mathbf{C}_t^{(1)}$ ,  $\alpha$ ,  $R$  and  $\mathbf{A}$  we get the pdf of  $\mathbf{C}_t^{(1)}$  and  $\mathbf{A}$  given by

$$\begin{aligned}
 h(\mathbf{c}_t^{(1)}, \boldsymbol{\lambda}) &\propto |\boldsymbol{\Sigma}_{t11}|^{-\frac{1}{2}} |\mathbf{A}_t^{(1)T} \boldsymbol{\Sigma}_{t11}^{-1} \mathbf{A}_t^{(1)}|^{-\frac{1}{2}} \\
 &\quad \times \left( a_0 + \sum_{l=1}^s a_l \lambda_l + \mathbf{c}_t^{(1)T} \mathbf{K}_t \mathbf{c}_t^{(1)} \right)^{\frac{1}{2}(n_t + \sum_{l=0}^s h_l - p)} \prod_{l=1}^s \lambda_l^{\frac{1}{2}h_l - 1}
 \end{aligned} \tag{15}$$

So the conditional pdf of  $\mathbf{C}_t^{(2)}$  given  $\mathbf{A} = \boldsymbol{\lambda}$  is given by

$$h(\mathbf{c}_t^{(2)} | \mathbf{C}_t^{(1)} = \mathbf{c}_t^{(1)}, \mathbf{A} = \boldsymbol{\lambda})$$

$$\begin{aligned}
& \propto \frac{f(\mathbf{c}_t, \boldsymbol{\lambda})}{h(\mathbf{c}_t^{(1)}, \boldsymbol{\lambda})} \\
& \propto \left( \frac{|\boldsymbol{\Sigma}_t|}{|\boldsymbol{\Sigma}_{t11}|} \right)^{-\frac{1}{2}} \left( \frac{|\mathbf{A}_t^T \boldsymbol{\Sigma}_t^{-1} \mathbf{A}_t|}{|\mathbf{A}_t^{(1)T} \boldsymbol{\Sigma}_{t11}^{-1} \mathbf{A}_t^{(1)}|} \right)^{-\frac{1}{2}} \\
& \times \left( a_0 + \sum_{l=1}^s a_l \lambda_l + \mathbf{c}_t^{(1)T} \mathbf{K}_t \mathbf{c}_t^{(1)} \right)^{\frac{1}{2}(n_t + \sum_{l=0}^s h_l - p)} \\
& \times \left( a_0 + \sum_{l=1}^s a_l \lambda_l + \mathbf{c}_t^T \mathbf{Q}_t \mathbf{c}_t \right)^{\frac{1}{2}(N_t + \sum_{l=0}^s h_l - p)} \quad (16)
\end{aligned}$$

From (7)

$$\begin{aligned}
|\mathbf{A}_t^T \boldsymbol{\Sigma}_t^{-1} \mathbf{A}_t| &= |\mathbf{A}_t^{(1)T} \boldsymbol{\Sigma}_{t11}^{-1} \mathbf{A}_t^{(1)} + (\mathbf{A}_t^{(2)} - \boldsymbol{\Sigma}_{t21} \boldsymbol{\Sigma}_{t11}^{-1} \mathbf{A}_t^{(1)})^T \boldsymbol{\Sigma}_{t22.1}^{-1} \\
&\quad \times (\mathbf{A}_t^{(2)} - \boldsymbol{\Sigma}_{t21} \boldsymbol{\Sigma}_{t11}^{-1} \mathbf{A}_t^{(1)})| \\
&= \left| \begin{array}{cc} \mathbf{A}_t^{(1)T} \boldsymbol{\Sigma}_{t11}^{-1} \mathbf{A}_t^{(1)} & -(\mathbf{A}_t^{(2)} - \boldsymbol{\Sigma}_{t21} \boldsymbol{\Sigma}_{t11}^{-1} \mathbf{A}_t^{(1)})^T \\ \mathbf{A}_t^{(2)} - \boldsymbol{\Sigma}_{t21} \boldsymbol{\Sigma}_{t11}^{-1} \mathbf{A}_t^{(1)} & \boldsymbol{\Sigma}_{t22.1} \end{array} \right| \\
&\quad \div |\boldsymbol{\Sigma}_{t22.1}| \\
&= \left| \mathbf{A}_t^{(1)T} \boldsymbol{\Sigma}_{t11}^{-1} \mathbf{A}_t^{(1)} \right| \left| \boldsymbol{\Sigma}_{t22.1} + (\mathbf{A}_t^{(2)} - \boldsymbol{\Sigma}_{t21} \boldsymbol{\Sigma}_{t11}^{-1} \mathbf{A}_t^{(1)})^T \right. \\
&\quad \times (\mathbf{A}_t^{(1)T} \boldsymbol{\Sigma}_{t11}^{-1} \mathbf{A}_t^{(1)})^{-1} (\mathbf{A}_t^{(2)} - \boldsymbol{\Sigma}_{t21} \boldsymbol{\Sigma}_{t11}^{-1} \mathbf{A}_t^{(1)}) \left. \right| \\
&\quad \div |\boldsymbol{\Sigma}_{t22.1}| \quad (17)
\end{aligned}$$

by using Exercise 2.4 on page 32 of Rao (1973) twice.

Also

$$|\boldsymbol{\Sigma}_t| = |\boldsymbol{\Sigma}_{t11}| |\boldsymbol{\Sigma}_{t22.1}| \quad (18)$$

Using the definition of  $\mathbf{G}_t$ , (17) and (18), we have

$$|\mathbf{A}_t^T \boldsymbol{\Sigma}_t^{-1} \mathbf{A}_t| = \left| \mathbf{A}_t^{(1)T} \boldsymbol{\Sigma}_{t11}^{-1} \mathbf{A}_t^{(1)} \right| |\mathbf{G}_t| \div \left( \frac{|\boldsymbol{\Sigma}_t|}{|\boldsymbol{\Sigma}_{t11}|} \right)$$

$$\left( \frac{|A_t^T \Sigma_t^{-1} A_t|}{|A_t^{(1)T} \Sigma_{t11}^{-1} A_t^{(1)}|} \right) \left( \frac{|\Sigma_t|}{|\Sigma_{t11}|} \right) = |G_t| \quad (19)$$

So the conditional pdf of  $C_t^{(2)}$  given  $C_t^{(1)}$  and  $\lambda$  is given by

$$\begin{aligned} h(c_t^{(2)} | c_t^{(1)}, \lambda) &\propto |G_t|^{-\frac{1}{2}} \left( a_0 + \sum_{l=1}^s a_l \lambda_l + c_t^{(1)T} K_t c_t^{(1)} \right)^{\frac{1}{2}(N_t - n_t)} \\ &\times [1 + \left( a_0 + \sum_{l=1}^s a_l \lambda_l + c_t^{(1)T} K_t c_t^{(1)} \right)^{-1} (c_t^{(2)} - M_t c_t^{(1)})^T \\ &\times G_t^{-1} (c_t^{(2)} - M_t c_t^{(1)})]^{-\frac{1}{2}(N_t + \sum_{l=0}^s h_l - p)} \end{aligned} \quad (20)$$

Writing

$$\left( N_t + \sum_{l=0}^s h_l - p \right)^{-1} \left( a_0 + \sum_{l=1}^s a_l \lambda_l + c_t^{(1)T} K_t c_t^{(1)} \right) G_t = \Omega_t$$

We have

$$\begin{aligned} f(c_t^{(2)} | c_t^{(1)}, \lambda) &\propto |\Omega_t|^{-\frac{1}{2}} \left\{ n_t + \sum_{l=0}^s h_l - p + (c_t^{(2)} - M_t c_t^{(1)})^T \Omega_t^{-1} \right. \\ &\times [(c_t^{(2)} - M_t c_t^{(1)})]^{-\frac{1}{2}(N_t + \sum_{l=0}^s h_l - p)} \} \end{aligned} \quad (21)$$

From (21) it follows that the conditional distribution of  $C_t^{(2)}$  given  $C_t^{(1)} = c_t^{(1)}$  and  $\Lambda = \lambda$  is  $(N_t - n_t)$ -variate  $t$  distribution with df  $n_t + \sum_{l=0}^s h_l - p$ , location parameter  $M_t c_t^{(1)}$  and scale parameter

$$\Omega_t = (N_t + \sum_{l=0}^s h_l - p)^{-1} \left( a_0 + \sum_{l=1}^s a_l \lambda_l + c_t^{(1)T} K_t c_t^{(1)} \right) G_t$$

Also, since  $f(\lambda | c_t^{(1)}) = \frac{f(c_t^{(1)}; \lambda)}{f(c_t^{(1)})} \propto h(c_t^{(1)}, \lambda)$ , we have from (15)

$$f(\lambda | c_t^{(1)}) \propto |\Sigma_{t11}|^{-\frac{1}{2}} |A_t^{(1)T} \Sigma_{t11}^{-1} A_t^{(1)}|^{-\frac{1}{2}} \prod_{l=1}^s \lambda_l^{\frac{1}{2}h_l - 1}$$

$$\times \left( a_0 + \sum_{l=1}^s a_l \lambda_l + \mathbf{c}_t^{(1)T} \mathbf{K}_t \mathbf{c}_t^{(1)} \right)^{-\frac{1}{2}(n_t + \sum_{l=0}^s h_l - p)}$$



## BIBLIOGRAPHY

- Battese, G.E., Harter, R.M., and Fuller, W.A. (1988). An error-components model for prediction of county crop areas using survey and satellite data. *J. Amer. Statist. Assoc.*, 83, 28-36.
- Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion). *J. R. Statist. Soc.B* 36, 192-326.
- Blight B.J.N. and Scott A.J. (1973). A stochastic model for repeated surveys. *J. R. Statist. Soc.B* 35, 61-66.
- Brackstone, G.J. (1987). Small area data: policy issues and technical challenges. *Small Area Statistics*. Eds. R. Platek, J.N.K., Rao, C.E., Sarndal, and M.P. Singh. Wiley, New York, pp. 3-20.
- Datta, G.S. (1990). Bayesian prediction in mixed linear models with applications in small area estimation. Unpublished Ph.D. dissertation. Department of Statistics, University of Florida.
- Datta, G.S., Fay, R.E. and Ghosh, M. (1991). Hierarchical and empirical multivariate Bayes analysis in small area estimation. Technical Report 409, Dept. Statistics, Univ. Florida.
- Datta, G.S. and Ghosh, M. (1988). Minimum risk equivariant estimators of percentiles in location-scale families of distributions. *Cal. Statist. Assoc. Bull.*, 37, 201-207.
- Datta, G.S. and Ghosh, M. (1990). On the frequentist properties of some hierarchical Bayes predictors in finite population sampling. Technical Report, Dept. Statistics, Univ. Florida.
- Datta, G.S. and Ghosh, M. (1991). Bayesian prediction in linear models: Applications to small area estimation. *The Annals of Statistics*, 19, 1748-1770.
- Efron, B., and Morris, C. (1971). Limiting the risk of Bayes and empirical Bayes estimators—Part I: the Bayes case. *J. Amer. Statist. Assoc.*, 66, 807-815.
- Ericksen, E.P. (1973). A method of combining sample survey data and symptomatic indicators to obtain population estimates for local areas. *Demography*, 10, 137-160.

- Ericksen, E.P. (1974). A regression method for estimating populations of local areas. *J. Amer. Statist. Assoc.*, 69, 867-875.
- Ericksen, E.P. and Kadane, J.B. (1985). Estimating the population in a census year (with discussion). *J. Amer. Statist. Assoc.*, 80, 98-131.
- Ericksen, E.P. and Kadane, J.B. (1987). Sensitivity analysis of local estimates of undercount in the 1980 U.S. Census. *Small Area Statistics*. Eds. R. Platek, J.N.K., Rao, C.E., Sarndal, and M.P. Singh. Wiley, New York, pp. 23-45.
- Ericksen, E.P., Kadane, J.B., and Tukey, J.W. (1989). Adjusting the 1980 Census of population and housing. *J. Amer. Statist. Assoc.*, 84, 927-944.
- Fay, R.E. (1987). Application of multivariate regression to small domain estimation. *Small Area Statistics*. Eds. R. Platek, J.N.K., Rao, C.E., Sarndal, and M.P. Singh. Wiley, New York, pp. 91-102.
- Fay, R.E., and Herriot, R.A. (1979). Estimates of income for small places: an application of James-Stein procedures to census data. *J. Amer. Statist. Assoc.*, 74, 269-277.
- Gelfand, A.E. and Smith, A.F.M. (1990). Sampling based approaches to calculating marginal densities. *J. Amer. Statist. Assoc.*, 85, 398-409.
- Gelfand, A.E. and Smith, A.F.M. (1991). Gibbs Sampling for Marginal Posterior Expectations. *Commun. in Statist.-Theory Meth.*, 20(5 & 6), 1747-1766..
- Ghosh, M. and Lahiri, P. (1987). Robust empirical Bayes estimation of means from stratified samples. *J. Amer. Statist. Assoc.*, 82, 1153-1162.
- Ghosh, M. and Lahiri, P. (1988). A hierarchical Bayes approach to small area estimation with auxiliary information. *Bayesian analysis in statistics and econometrics*. Eds., P.K. Goel and N.S. Iyengar. Lecture notes in statistics, #75. Springer-Verlag, New York, pp 107-125.
- Ghosh, M. and Meeden, G. (1986). Empirical Bayes estimation in finite population sampling. *J. Amer. Statist. Assoc.*, 74, 269-277.
- Ghosh, M. and Rao, J.N.K. (1991). Small area estimation: An appraisal. Technical Report 390, Dept. Statistics, Univ. Florida.
- Hwang, J.T. (1985). Universal domination and stochastic domination: estimation simultaneously under a broad class of loss functions. *The Annals of Statistics*, 13, 295-314.
- Jessen, R.J. (1942). Statistical investigation of a sample survey for obtaining farm facts. *Iowa Agricultural Experimental Station, Research Bulletin*, 304, 54-59.

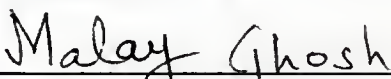
- Kelker, D. (1970). Distribution theory for spherical distribution and a location-scale parameter generalization. *Sankhya, A.*, **32**, 419-430.
- Lindley, D.V. and Smith, A.F.M. (1972). Bayes estimates for linear model (with discussion). *J. R. Statist. Soc.B* **34**, 1-41.
- Lehmann, D.V., and Scheffe, H. (1950). Completeness, similar regions and unbiased estimation. *Sankhya*, **10**, 305-340.
- Meinhold, R.J., and Singpurwalla, N.D. Understanding the Kalman filter. *The American Statistician*, **37**, 123-137.
- Patterson, D. (1950). Sampling on successive occasions with partial replacement of units. *J. R. Statist. Soc.B* **12**, 241-255.
- Prasad, N.G.N., and Rao, J.N.K. (1990). The estimation of mean squared errors of small-area estimators. *J. Amer. Statist. Assoc.*, **85**, 163-171.
- Raj, D. (1965). On sampling over two occasions with probability proportionate to size. *Annals of Math. Statist.*, **36**, 327-330.
- Rao, C.R. (1952). Some theorems on minimum variance estimation. *Sankhya*, **12**, 27-42.
- Rao, C.R. (1973). Linear Statistical Inference and its Applications. 2nd Edition. Wiley, New York.
- Rao, J.N.K. and Graham, J.E. (1964). Rotation designs for sampling on repeated occasions. *J. Amer. Statist. Assoc.*, **59**, 492-509.
- Rodrigues, J. and Bolfarine, H. (1987). A Kalman filter model for single and two-stage repeated surveys. *Statist. and Probab. Letters.*, **5**, 299-303.
- Sallas, W.M. and Harville, D.A. Rodrigues, J. (1988). Best linear recursive estimation for mixed linear models. *J. Amer. Statist. Assoc.*, **76**, 860-869.
- Scott A.J. and Smith T.M.F. (1974). Analysis of repeated surveys using time series models. *J. Amer. Statist. Assoc.*, **69**, 674-678.
- Singh, D. (1968). Estimates in successive sampling using a multi-stage design. *J. Amer. Statist. Assoc.*, **63**, 99-112.

## BIOGRAPHICAL SKETCH

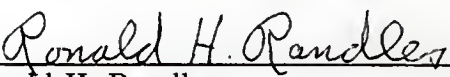
Narinder K. Nangia was born on May 15, 1955, in New Delhi, India. He received his Bachelor of Science degree with honors in mathematics in 1975 and thereafter master's degree in mathematical statistics in 1977, both from the University of Delhi, India. In 1979, he joined the Indian Statistical Service of the Government of India. As a consequence, he received training in various fields of statistics, economic administration and planning during 1979–1981 in different institutions including the Indian Statistical Institute. On successful completion of his training in 1981, he joined the Bureau of Industrial Costs and Prices as an Assistant Director. Later, in 1983 he was promoted to the rank of Deputy Director in the same institution. In 1987, he was posted in National Sample Survey Organisation. In August 1987, he joined the doctoral program in statistics in the University of Florida at Gainesville. He expects to receive a Ph.D. degree in December 1992.



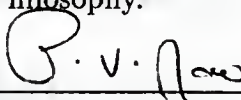
I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

  
\_\_\_\_\_  
Malay Ghosh, Chairman  
Professor of Statistics

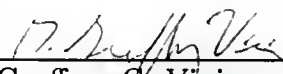
I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

  
\_\_\_\_\_  
Ronald H. Randles  
Professor of Statistics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

  
\_\_\_\_\_  
Pejaver V. Rao  
Professor of Statistics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

  
\_\_\_\_\_  
Geoffrey G. Vining  
Assistant Professor of Statistics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

  
\_\_\_\_\_  
Sencer Yeralan  
Associate Professor of Industrial  
and Systems Engineering



This dissertation was submitted to the Graduate Faculty of the Department of Statistics in the College of Liberal Arts and Sciences and to the Graduate School and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

December 1992

---

Dean, Graduate School

UNIVERSITY OF FLORIDA



3 1262 08554 0432